



Dynamiques multi-échelle de l'ADN-B

Akli Ben Imeddourene

► **To cite this version:**

Akli Ben Imeddourene. Dynamiques multi-échelle de l'ADN-B. Biologie structurale [q-bio.BM]. Université Paris-Saclay, 2015. Français. <NNT : 2015SACLN025>. <tel-01325227>

HAL Id: tel-01325227

<https://tel.archives-ouvertes.fr/tel-01325227>

Submitted on 2 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



université
PARIS-SACLAY



2015SACLN025

UNIVERSITÉ PARIS SACLAY

ÉCOLE DOCTORALE STRUCTURE ET DYNAMIQUE DES SYSTEMES VIVANTS

Spécialité de doctorat : Sciences de la vie et de la santé

Présenté par

Akli BEN IMEDDOURENE

Pour obtenir le grade de

Docteur de l'Université Paris Saclay

Dynamique multi-échelle de l'ADN-B

Soutenue le 21 décembre 2015 devant le jury composé de :

**Pr Annick DEJAEGERE
Dr Carine VAN-HEIJENOORT**

**Dr Carine TISNÉ
Pr Catherine ETCHEBEST**

**Dr Brigitte HARTMANN
Dr Olivier MAUFFRET**

**Rapporteur
Rapporteur**

**Examinatrice
Présidente**

**Directrice de thèse
Directeur de thèse**

Remerciements

En préambule de ce manuscrit, je souhaite adresser mes remerciements les plus sincères aux personnes qui m'ont apporté leur aide et leur soutien et qui ont ainsi contribué de près ou de loin à l'élaboration de ce travail.

Mes travaux de Thèse ont été effectués au Laboratoire de Biologie et de Pharmacologie Appliquée (LBPA) à l'École Normale Supérieure de Cachan. Je tiens à remercier le Directeur du LBPA, Malcolm Buckle, de m'avoir accueilli et pour son soutien aussi fort que constant durant toutes les années que j'ai passé dans son laboratoire.

Je remercie le Pr. Annick Dejaegere, rapporteur de mon jury de thèse, pour avoir lu attentivement mon manuscrit et pour son regard critique sur mon travail. Je tiens à remercier chaleureusement le Dr. Carine Van-Heijenoort, pour son acceptation de rapporter ma thèse, et, plus particulièrement, pour toutes les discussions scientifiques que nous avons eues, et qui m'ont permis d'améliorer mon manuscrit. De même, je remercie vivement le Pr. Catherine Etchebest et le Dr. Carine Tisné d'avoir accepté d'examiner mes travaux.

Ma thèse a été réalisée sous la direction du Dr. Brigitte Hartmann et du Dr. Olivier Mauffret. Je souhaiterais leur exprimer ma gratitude de m'avoir intégré dans l'équipe. Ils m'ont soutenu depuis mes premiers pas dans un laboratoire de recherche jusqu'à l'aboutissement de mon doctorat. En travaillant sous leur direction, j'ai bénéficié de leur grande expérience dans le domaine de la biologie structurale et de toutes leurs compétences en RMN et en modélisation moléculaire. J'espère avoir su en retirer une formation très complète. Merci aussi de m'avoir laissé expérimenter librement mes propres idées et, par de très nombreuses discussions scientifiques et techniques, de m'avoir guidé pour me perfectionner.

Mes remerciements les plus sincères s'adressent tout particulièrement au Dr. Loussiné Zargarian pour son initiation aux techniques RMN, sa présence à mes côtés durant toutes ces années, et son implication sans faille. Je lui dois d'avoir acquis une maîtrise et une autonomie suffisante pour mener à bien les expériences que j'ai entrepris en dehors du laboratoire. Merci infiniment pour tous ses conseils et tout le temps qu'elle m'a consacré.

Je tiens également à remercier les étudiants avec qui j'ai travaillé et échangé. À commencer par Ahmad Elbhansi, dont j'ai partagé le bureau depuis son arrivée au laboratoire. Merci pour son aide, si précieuse pour mes travaux de modélisation moléculaire. Je souhaiterais aussi exprimer mes remerciements à Mademoiselle Xiaoxian Xu qui a initié avant moi les travaux RMN de mon projet de thèse.

Je tiens à remercier les chercheurs qui ont collaboré à mon sujet. Merci à Christophe Oguey pour toutes ses analyses statiques, sans lesquelles l'interprétation des résultats aurait été incomplète. Je tiens également à remercier Nicolas Foloppe, pour ses conseils ingénieux et son implication. Je tiens également à exprimer ma reconnaissance à Claude Noges, pour avoir partagé les données expérimentales obtenues dans son équipe et que j'ai pu présenter lors de ma soutenance orale.

Durant ma thèse j'ai pu utiliser les appareils RMN de l'ICSN à Gif-sur-Yvette, ce qui m'a permis de réduire considérablement le temps des expériences. Je tiens donc à adresser mes remerciements à toute l'équipe RMN de l'ICSN. Un grand merci à Nelly Morrelet, avec qui j'ai travaillé directement et qui m'a consacré tout le temps nécessaire pour mener à bien mes expériences. Je remercie également Ewen Lescop pour toutes les discussions

techniques que nous avons échangées. Merci aussi à Christina, Nadine, Sevrine et à tous les autres membres du laboratoire dont j'ai partagé le déjeuner ou le café. Je tiens aussi à remercier les ingénieurs de Super Calculateur "Genci" pour leur réactivité et leur aide.

J'ai pu mener mes travaux de thèse dans un climat agréable et chaleureux au sein de l'équipe Structures et Interactions des acides nucléiques. Je tiens donc à remercier vivement l'ensemble des membres de l'équipe, étudiants et permanents, Ouerdia Maskri, Anissa Belfatmi, Cel Areny-Navès, Brigitte René, Françoise Chaminade, Élisabeth Bugnard, Philippe Fossé et Xu-Guang Xi, ainsi que toutes les personnes du LBPA que j'ai côtoyées.

Enfin, je tiens à remercier ma famille pour son soutien pendant toutes ces années.

Table des matières

INTRODUCTION.....	1
1 Contexte général.....	1
2 Variabilité conformationnelle dans les structures cristallographique d'ADN-B.....	3
2.1 Les biais des structures cristallographiques.....	3
2.2 Mise en évidence de la mécanique de l'ADN-B et probable effet de séquence.....	6
3 L'étude des mouvements rapides de l'ADN par RMN en solution.....	11
3.1 Principes des approches RMN permettant de sonder directement les mouvements.....	11
3.1.1 Les NOE hétéronucléaires et la relaxation.....	11
3.1.2 Le paramètre d'ordre S^2	12
3.2 Résultats obtenus par relaxation des spins et calculs des paramètres d'ordre S^2	14
4 Les mouvements des groupes phosphate dans l'ADN-B en RMN.....	16
4.1 Mise en évidence et quantification de l'équilibre BI \leftrightarrow BII.....	16
4.2 Effet de la séquence de l'ADN sur la dynamique BI \leftrightarrow BII	20
5 Simulations de la dynamique des ADN à l'échelle de pico- nanoseconde.....	24
5.1 Le cas particulier des affinements des structures d'ADN à partir de données RMN.....	25
5.2 Dynamiques moléculaires et champs de force empiriques.....	27
6 L'étude des mouvements lents dans l'ADN par RMN en solution.....	33
6.1 Introduction à la relaxation des spins.....	33
6.2 L'échange conformationnel et la relaxation des spins.....	34
6.2.1 Théorie générale de l'échange conformationnel (" <i>chemical exchange</i> ") dans le cas d'un système à deux états.....	34
6.2.2 L'échange rapide ($k_{ex} > \Delta w$)	36
6.2.3 L'échange lent ($k_{ex} < \Delta w$)	37
6.3 Stratégie d'étude des mouvements lents des ADN.....	38
6.3.1 Les expériences RMN de mesure des vitesses de relaxation.....	39
6.3.2 Expériences de dispersion relaxation.....	42
6.4 Les échanges conformationnels mis en évidence au sein des ADN-B.....	51
7 Apport personnel.....	53
CHAPITRE 1 : L'échelle TRX et la formation du nucléosome.....	55
1 Introduction.....	55
2 Article 1.....	57
CHAPITRE 2: Test de champs de force et nouveaux aspects de la dynamique des groupements phosphates	76
1 Introduction.....	76
2 Article 2.....	80
CHAPITRE 3 : Mécanique de l'ADN-B en solution.....	131
1 Introduction.....	131

2 Article 3.....	136
CHAPITRE 4 : Mouvements lents de l'ADN-B par RMN.....	169
1 Introduction générale.....	169
2 Méthodologie.....	170
2.1 Préparation de l'échantillon.....	170
2.2 Expériences RMN.....	170
2.2.1 Mesure de température de fusion.....	170
2.2.2 Mesure des largeurs des raies des protons H2.....	170
2.2.3 HSQC (" <i>Hétéronuclear Single-Quantum Correlation</i> ") à temps constant.....	171
2.2.4 Calibration de la force de champ de verrouillage de spin (" <i>spin lock</i> ") (ω_1) pour les expériences de dispersion de relaxation de type R1 ρ	171
2.2.5 Calcul de la contribution du Hartmann-Hahn à la relaxation.....	172
2.2.6 Expériences 1D " <i>On resonance</i> " R1 ρ	175
2.3 Spectromètres.....	175
2.4 Traitement et analyse des données	175
3 Résultats et discussion.....	178
3.1 Mesure de la largeur à mi-hauteur des résonances des protons H2 des adénines.....	178
3.2 Dispersion relaxation ^{13}C	183
3.2.1 Calibration de la force de champ	183
3.2.2 Calcul de la contribution de Hartmann-Hahn à la relaxation.....	185
3.2.3 Dispersion relaxation " <i>On Resonance</i> " R1 ρ	188
4 Conclusion.....	197
CONCLUSION ET PERSPECTIVES.....	199
1 Mouvements rapides.....	199
2 Mouvements lents.....	201
ANNEXE A.....	203
ANNEXE B.....	204
BIBLIOGRAPHIE.....	205

INTRODUCTION

1 Contexte général

L'acide désoxyribonucléique (ADN) est la molécule support de l'information génétique. Chez les organismes eucaryotes l'ADN a besoin d'être compacté pour permettre le maintien de près de deux mètres d'ADN dans un noyau de quelques micromètres mais aussi d'être décompacté pour être transcrit, répliqué, réparé, et enfin dégradé à la fin de la vie de la cellule. L'interaction de l'ADN avec les protéines est le socle de ces nombreuses fonctions.

Notre laboratoire s'intéresse aux interactions ADN-protéines et plus particulièrement à la compaction de l'ADN, dont l'unité de base dans les cellules eucaryotes est le nucléosome. Le nucléosome est constitué de 8 protéines histones autour desquelles 145 à 147 paires de bases d'ADN sont enroulées. Ce type de complexe est qualifié de non spécifique puisqu'il doit se former avec n'importe quelle séquence d'ADN. En fait, on sait que le positionnement des nucléosomes le long des génomes n'est pas homogène, ce qui a suggéré l'idée que certaines séquences d'ADN auraient des propriétés qui favoriseraient la formation des nucléosomes. L'enroulement de l'ADN autour du cœur protéique provoque en effet de fortes distorsions de la double hélice et l'on s'attend à ce que la formation et la stabilité du complexe soit reliée aux propriétés de malléabilité intrinsèque de l'ADN, dépendantes de la séquence, qui vont moduler la capacité de l'ADN à s'adapter au cœur d'histone. Cette hypothèse est supportée par l'identification par la méthode SELEXION (*Systematic Evolution of Ligands by EXponential enrichment with Integrated Optimization by Non-linear analysis*) (Irvine, Tuerk, et Gold 1991) d'une série de séquences d'ADN synthétiques (non naturel) qui ont une forte affinité pour le cœur d'histone (Thåström, Bingham, et Widom 2004). La séquence la plus apte à former des nucléosome est appelée «séquence 601». Cette séquence a été étudiée par différentes méthodes biophysiques (Pisano et al. 2006, Mihardja et al. 2006, Morozov et al. 2009) et il existe trois structures cristallographiques de nucléosome la contenant (Makde et al. 2010) (Vasudevan, Chua, et Davey 2010). Mais malgré de nombreux efforts, les propriétés de l'ADN favorisant ou défavorisant la formation du nucléosome restent insaisissables. La seule certitude est qu'une alternance de régions riches en A:T et riches en G:C caractérise les séquences de plus grande affinité pour le cœur d'histone.

Ce type de question concerne non seulement le nucléosome mais beaucoup plus largement l'ensemble des interactions de l'ADN avec des protéines, spécifiques ou non. Intuitivement, il est clair que les propriétés de malléabilité de l'ADN vont jouer un rôle important dans ces interactions quand il s'agira d'assurer la complémentarité structurale avec les partenaires protéiques. Un des objectifs en biologie structurale est donc de décrire cette malléabilité intrinsèque et de savoir comment elle peut varier en fonction de la séquence. Au vu du grand nombre de séquences cibles de protéine, il faudrait idéalement pouvoir prédire ces propriétés. Ces thématiques sont au cœur de mon sujet de thèse.

Tout d'abord, il est important de préciser ce que nous entendons par malléabilité intrinsèque de l'ADN, un terme qui peut correspondre à plusieurs interprétations. Nous définirons donc ici la malléabilité intrinsèque de l'ADN comme l'espace conformationnel exploré spontanément (sans contrainte extérieure) par l'ADN « nu » (sans protéine) en conformation B, qui reflète la dynamique de la double hélice à l'échelle atomique.

De fait, connaître cet espace conformationnel et décrypter l'effet de la séquence sont des tâches extrêmement difficiles à réaliser, parce que les variations autour de la structure « canonique » sont très fines. Cette difficulté sera illustrée dans l'introduction qui suit et qui présente l'état de nos connaissances dans le domaine de la dynamique des ADN, impliquant soit des mouvements rapides, à l'échelle de la pico-nanoseconde, soit des mouvements plus lents, à l'échelle de la micro-milliseconde. Nous commencerons par détailler les travaux réalisés sur la variabilité de structure à partir de données cristallographiques. Les structures cristallographiques étant statiques, cette variabilité n'est pas à proprement parler une approche de la dynamique de l'ADN mais elle donne une bonne idée du paysage conformationnel accessible. Nous présenterons ensuite les avancées obtenues par Résonance Magnétique Nucléaire (RMN) et par modélisation moléculaire, deux techniques qui s'adressent à des ADN en solution.

2 Variabilité conformationnelle dans les structures cristallographique d'ADN-B

La première structure cristallographique à haute résolution (1.9 Å) d'un dodécamère en conformation B en 1980 par le groupe de Dickerson (Wing et al. 1980) a ouvert une nouvelle voie

pour les études structurales des ADN. Depuis cette époque, le nombre de structures a augmenté et permet de faire des analyses statistiques sur la variabilité conformationnelle des différents composants des ADN, décrits par les angles du squelette phosphodiester, les phases des sucres et les paramètres hélicoïdaux. Néanmoins, ce type d'approche est confronté à des limitations inhérentes aux structures d'ADN-B cristallisées que nous allons d'abord détailler.

2.1 Les biais des structures cristallographiques

Premièrement, il ne faut pas oublier que les conditions de cristallisation sont assez éloignées des conditions physiologiques, notamment avec une forte concentration en facteurs de cristallisation non biologiques, et les températures sont souvent très basses pour préserver les cristaux.

Un autre problème vient du fait que les détails fins des structures tridimensionnelles, comme notamment les angles de torsion du squelette phosphodiester, peuvent dépendre du protocole de raffinement des clichés de diffraction. Ainsi, les modèles issus de différents algorithmes de raffinement appliqués au même cliché de diffraction résolu à 3.4 Å montrent d'importantes variations de structure (Holbrook, Dickerson, et Kim 1985). Ce type de biais est limité si la résolution est de moins de 2Å (Heinemann, Alings, et Hahn 1994), les structures récemment obtenues à haute ou très haute (<1Å) résolution étant naturellement les plus précises. Toutefois, les biais peuvent aussi provenir de l'effet de la maille cristalline sur la structure de l'ADN à travers des contacts intermoléculaires ADN-ADN, ainsi que l'a montré la résolution d'une même séquence d'ADN dans trois groupes d'espaces différents (Shakked et al. 1989). Des travaux ultérieurs (Heinemann, Alings, et Hahn 1994)(Heinemann, Alings, et Hahn 1994) ont montré que deux décimères de séquence CCAACGTTGG et CCAACITTGG résolus dans le même groupe d'espace C2 présentaient un écart de structure faible (Root Mean Square Deviation RMSD- de 0.5Å) du à la mutation de la guanine par une inosine (en gras dans les séquences) tandis que le RMSD de la même séquence résolue dans deux groupes d'espace différents était plus élevé (RMSD de 2.5 Å). L'analyse de 60 structures cristallographiques d'ADN-B a permis de définir trois principales familles d'empilement dans un environnement cristallin (Figure 1) (Berman 1997).

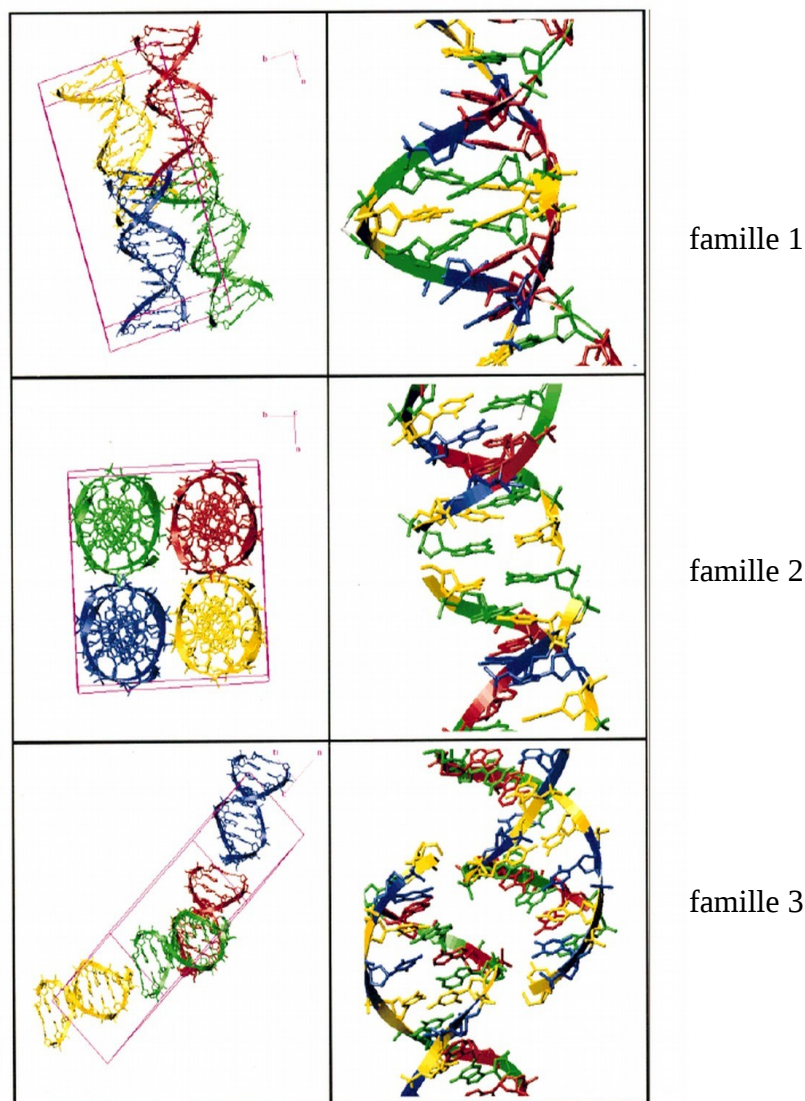


Figure 1: Exemple des trois principales familles de compaction d'ADN-B dans les mailles cristallines. Les modèles simplifiés de mailles cristallines sont illustrés à gauche, avec quatre couleurs pour distinguer quatre molécules dans l'unité asymétrique (en magenta). Une vue plus détaillée des motifs de liaison est donnée à droite (adénines en rouge, thymine en bleu, guanines en vert et cytosines en jaune). La première famille correspond à des dodécamères qui interagissent par leur petit sillon (code NDB : BDL042). La deuxième montre des contacts latéraux entre décimères (code NDB : BDJ025) et la troisième représente les interactions entre grand sillon et squelette phosphodiester, également pour des décimères (code NDB : BDJ060). D'après (Berman 1997)

La famille 2, spécifique des décimères, correspond aux contacts ADN-ADN les moins susceptibles de perturber les structures (Berman 1997). Il existe bien des contacts latéraux entre deux molécules d'ADN, mais les bases terminales des ADN sont empilées en reconstruisant en quelque sorte un seul ADN continu. Les décimères appartenant à cette famille et résolus à moins de 2 Å s'avèrent donc correspondre au meilleur choix pour minimiser l'effet de l'environnement cristallin. Le Tableau 1 donne la liste de ces décimères, réalisée début 2015. Cependant, cet inventaire montre un autre

biais, qui concerne les séquences. Ces séquences comportent ainsi majoritairement des paires de base G:C en 1^{ière}, 2^{ième}, 9^{ième} et 10^{ième} positions. Peut-être plus gênant, certains dinucléotides sont sous-représentés : les pas ApC•GpT et TpA•TpA ne sont représentés que par 14 et 20 paires respectivement par rapport aux pas GpG•CpC et les ApA•TpT qui sont présents respectivement en 56 et 39 exemplaires.

Code NDB	Code PDB	R (Å)	Séquence
bd0114	3GGI	0.98	CCAGGCCTGG
bd0113	3GGB	0.96	CCAGGCCTGG
bd0115	3GGK	0.87	CCAGGCCTGG
bd0034	1EN8	0.95	CCAACGTTGG
bd0035	1EN9	0.95	CCAGCGCTGG
bd0036	1ENE	0.95	CCAGCGCTGG
bd0023	1D8G	0.74	CCAGTACTGG
bd0079	1ZF5	0.99	CCAGCGCTGG
bd0087	1ZFF	0.94	CCGAATTCGG
bd0033	1EN3	0.95	CCAACGTTGG
bd0080	1ZF7	1.05	CCGTCGACGG
bdj019	5DNB	1.40	CCAACGTTGG
bdj025	1D23	1.50	CGATCGATCG
bdj031	1D49	1.50	CGATTAATCG
bd0014	463D	1.45	GCGAATTCGC
bd0071	1SK5	1.15	CTTTTAAAAG
bd0018	476D	1.30	GCGAATTCGC
bd0081	1ZFB	1.65	CCGCCGGCGG
bd0077	1ZF0	1.50	CCGTTAACGG
bdj017	1BD1	1.60	CCAGGCCTGG
bdj052	158D	1.90	CCAAGCTTGG
bd0051	1IKK	1.60	CCTTTAAAGG
bd0082	1ZFC	2.00	CCGATATCGG
bdj037	1D57	2.00	CGATATATCG
bdj051	126D	2.00	CATGGCCATG
bdj036	1D56	1.70	CCAGCGCTGG
bd0032	1EHV	1.80	GCGAATTCGC
bdj060	196D	1.70	CTCTCGAGAG
bdj061	n.a.	1.95	CCACTAGTGG
bdj081	307D	1.85	CAAAGAAAAG
bd0019	477D	1.70	GCGAATTCGC
bd0073	1WQY	2.00	CCATTAATGG
bd0084	1ZFG	1.75	CCGAGCTCGG

Tableau 1 : Inventaire des structures cristallographiques de décamères d'ADN-B. Les codes NDB et PDB (n.a : la structure n'est pas déposée dans la PDB) de chaque ADN sont donnés dans les deux premières colonnes. R est la résolution. La séquence est donnée dans la dernière colonne.

Une dernière limitation pourrait provenir d'une sous-estimation de l'hétérogénéité des structures. Il a été montré récemment par une étude approfondie de clichés à très haute résolution (Maehigashi et

al. 2012) que deux conformations des groupements phosphate (BI et BII décrits plus loin en détail) pouvaient co-exister dans l'unité cristallographique. Les auteurs ont ainsi réinterprété quelques structures déjà publiées et mettent en garde contre les biais qui subsistent très probablement dans de nombreux autres ADN-B déposés dans la PDB.

Malgré ces biais, les structures cristallographiques ont été largement utilisées pour comprendre la mécanique intrinsèque de l'ADN et l'effet de la séquence sur la structure comme nous allons le présenter dans le paragraphe suivant.

2.2 Mise en évidence de la mécanique de l'ADN-B et probable effet de séquence

L'analyse de la première structure cristallographique par Dickerson et Drew (Drew et al. 1981) a précocement dévoilé que les paramètres hélicoïdaux *twist*, *tilt* et *roll* étaient variables le long d'un dodécamère. Cette structure a également montré que les sucres pouvaient adopter les conformations *Sud*, *Est* ou *Nord*. Parmi les cinq angles de torsion du squelette phosphodiester (Figures 2) $\alpha/\beta/\gamma$ adoptaient tous des conformations g-/trans/g+ tandis que ϵ/ζ étaient majoritairement g-/trans, mais adoptaient la conformation trans/g- sur deux pas CpG et GpC. Ces deux conformations ont ensuite été nommées BI ($\epsilon/\zeta = t/g-$) et BII ($\epsilon/\zeta = g/t$) et définies par le pseudo angle $\epsilon-\zeta$, négatif pour BI et positif pour BII (Fratini et al. 1982).

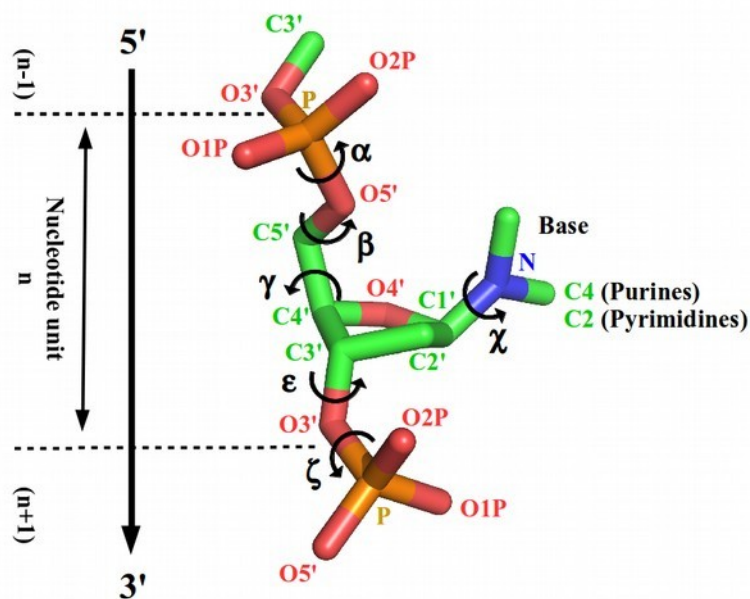


Figure 2: Illustration des différents angles de torsions dans l'ADN. Les angles du squelette phosphodiester sont α ($O3'_{n-1}-P-O5'-C5'$), β ($P-O5'-C5'-C4'$), γ ($O5'-C5'-C4'-C3'$), ϵ ($C4'-C3'-O3'-P$) et ζ ($C3'-O3'-P-O5'_{n+1}$); l'angle glycosylique χ ($O4'-C1'-N1-C2|C4$) définit l'orientation entre le sucre et la base aromatique.

Ces deux conformations correspondent à un basculement du groupe phosphate, montré dans la Figure 3.

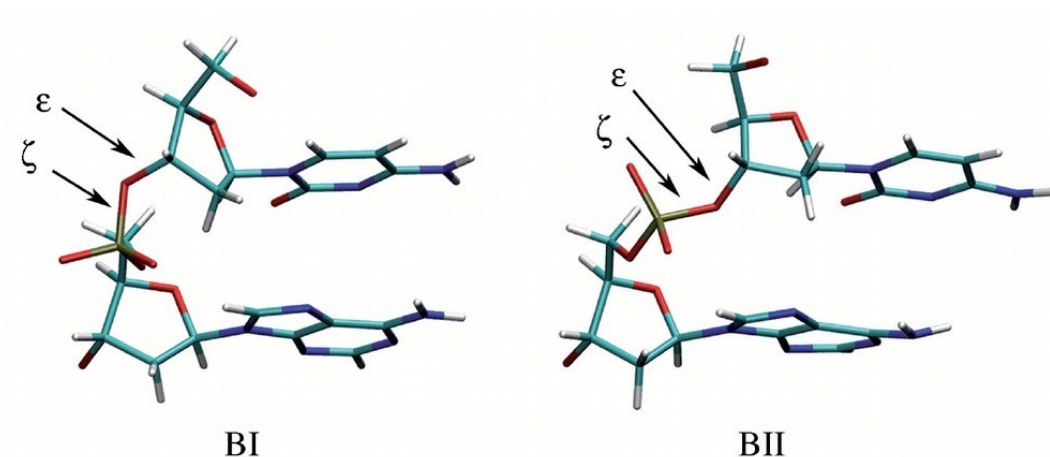


Figure 3: Illustration des états BI à gauche et BII à droite d'un pas CpA. La différence entre les deux états réside dans les angles de torsion ϵ et ζ , t/g- pour BI et g/t pour BII.

En 1991, le même groupe (Yanagi, Privé, et Dickerson 1991) publia les résultats des analyses des

structures de trois décimères et huit dodécimères qui contiennent les dix dinucléotides complémentaires constitutifs de l'ADN (AA.TT, GG.CC, GA.TC, AG.CT, AT.AT, AC.GT, GC.GC, TA.TA, CG.CG et CA.TG). Ces analyses ont mis en évidence les corrélations des paramètres hélicoïdaux *twist*, *roll* et *rise*, ainsi que l'effet de la séquence nucléotidique sur les paramètres hélicoïdaux. Les auteurs ont subdivisé les pas dinucléotidiques en deux catégories : i) les *high twist profile* (HTP), caractérisés par un grand *twist*, un *rise* faible et un *roll* négatif, spécifiques des pas GpC, GpA et les Y-CpA-R (Y:pyrimidine et R:purine) et ii) les *low twist profile* (LTP), caractérisés par un petit *twist*, un grand *rise* et un *roll* positif et représentés sur les pas CpG, GpG, ApG et les CpA autres que les Y-CpA-R.

Le groupe de Wilma Olson a ensuite étendu ce type d'analyse à un nombre beaucoup plus élevé de structures (38 dans (Gorin, Zhurkin, et K. 1995) et 34 dans (Olson et al. 1998)) résolues à moins de 3Å et en triant les pas dinucléotidiques en fonction de leur nature. Des corrélations *Twist-Roll*, *Tilt-Shift*, *Roll-Slide* et *Twist-Slide* ont ainsi été révélées. Néanmoins, les corrélations ne sont pas toutes de la même qualité (exemple dans la Figure 4). De fait, ces différences semblent reliées à la variabilité des paramètres, plus importantes pour certains pas que pour d'autres, ce qui suggère l'existence d'un effet de séquence.

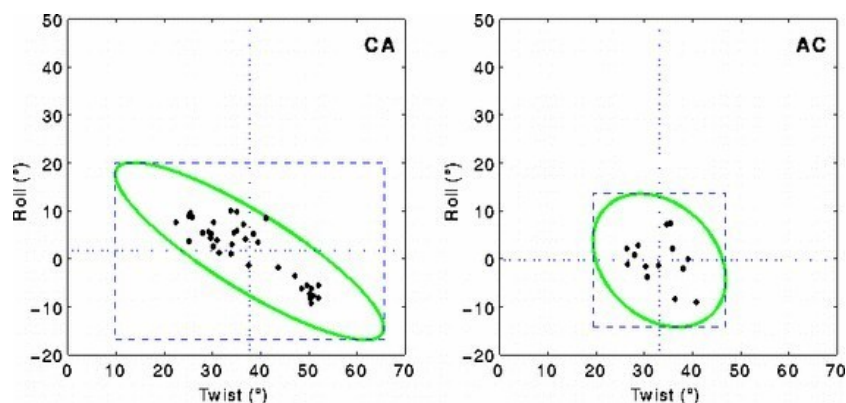


Figure 4 : Exemple de deux relations entre *Twist* et *Roll* dans des structures cristallographiques d'ADN-B. Source : <http://rutchem.rutgers.edu/~olson/pdna.html>.

L'étape suivante a été réalisée par le groupe de Brigitte Hartmann, qui a réintroduit dans les analyses les conformations des groupements phosphate et des sucres en plus des paramètres hélicoïdaux sur un ensemble de 57 structures à résolution de 2Å ou moins (Djuranovic et Hartmann 2003). Cette approche a permis de dévoiler une mécanique de l'ADN dans laquelle tous les composants des dinucléotides complémentaires sont solidaires les uns des autres, tel que résumé ci-

après.

Les phosphates ne peuvent adopter que deux conformations, BI et BII, alors que les angles α , β et γ sont invariants. Un phosphate en BII limite la variabilité des sucres adjacents, très majoritairement en *Sud* alors qu'ils peuvent adopter les conformations *Sud*, *Nord* ou *Est* quand les phosphates sont en BI. Pour les paramètres hélicoïdaux, décrits au niveau des dinucléotides complémentaires, il a été considéré les trois états possibles des phosphates qui se font face : BI.BI, BI.BII et BII.BII. Plus les pas sont riches en BII, plus le *Twist* augmente et plus le *Roll* devient négatif (Figure 5).

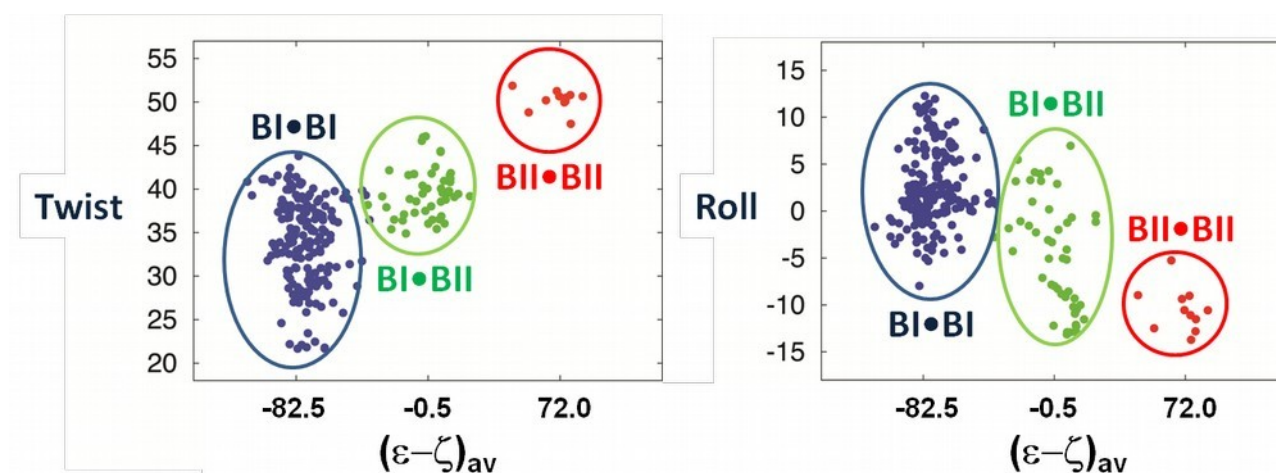


Figure 5 : Couplage au sein de dinucléotides complémentaires entre les états des deux phosphates face à face, le *Twist* (à gauche) et le *Roll* (à droite).

BI.BI, BI.BII et BII.BII sont également associés aux déplacements des paires de bases dans les sillons (*X-disp*), avec pour résultat un couplage avec les tailles des sillons. La densité de BII dans un tétramère (6 phosphates en tout) module en particulier la largeur du petit sillon (Figure 6) (Djuranovic et Hartmann 2004) , (Oguey, Foloppe, et Hartmann 2010).

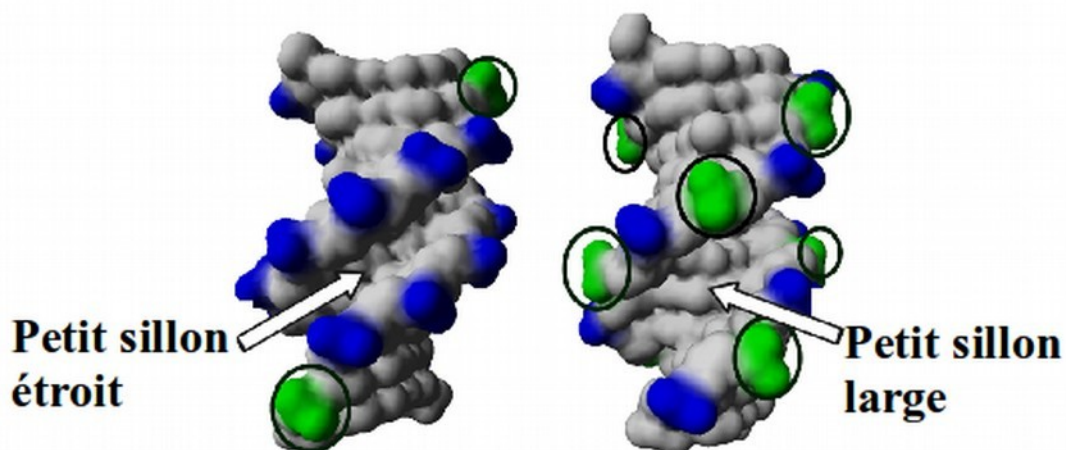


Figure 6 : Modulation de la largeur du petit sillon en fonction du nombre de phosphates en BII. La structure de gauche contient une large majorité de phosphates en BI (bleu) alors que dans la structure de droite un phosphate sur deux est en BII (vert).

Ces résultats ont mis en évidence une mécanique de l'ADN cohérente, dans laquelle les états conformationnels du phosphate jouent un rôle clé. Mais de plus, ils ont permis de montrer que les pourcentages de BII étaient dépendants de la séquence. Les phosphates des pas CpA.TpG et CpG.CpG paraissent ainsi particulièrement portés à adopter la conformation BII, alors que d'autres pas comme ApT.ApT ou ApA.TpT semblent réfractaires à cet état.

En conclusion, l'ensemble des analyses menées sur des structures cristallographiques a donc permis d'obtenir une vision assez précise de l'espace qui peut être exploré par les ADN-B. Elle a aussi apporté des éléments en faveur d'un effet de séquence sur la variabilité de la double hélice. Néanmoins il est difficile de mieux cerner ces points à cause des limitations exposées plus haut et évidemment du caractère statique des structures. Nous allons voir dans les chapitres suivant comment la RMN ou les simulations ont tenté d'avancer dans l'étude des mouvements rapides de l'ADN.

3 L'étude des mouvements rapides de l'ADN par RMN en solution

A priori la RMN en solution est la technique idéale pour appréhender la dynamique des macromolécules. Les expériences RMN les plus classiques permettent d'obtenir des distances inter et intra - nucléotidiques entre protons et des estimations d'angles dièdres. Néanmoins ces mesures reflètent une conformation moyenne et ne donnent pas d'informations directes sur la dynamique. Pour obtenir une vision des mouvements internes de l'ADN il faut réaliser des simulations de dynamique moléculaire sous contraintes expérimentales. Nous aborderons cette approche dans la partie 5.1 de cette introduction.

L'utilisation des NOE hétéronucléaires et le calcul de paramètres d'ordres S^2 ont par contre la capacité de sonder la dynamique rapide des ADN notamment au niveau des bases et des sucres. Dans ce chapitre, nous allons décrire brièvement les principes généraux de ces deux approches et décrire les résultats qui ont pu être obtenus en utilisant ces techniques

3.1 Principes des approches RMN permettant de sonder directement les mouvements

3.1.1 Les NOE hétéronucléaires et la relaxation

L'effet NOE (Nuclear Overhauser Effect) témoigne du couplage direct entre deux dipôles magnétiques. En l'absence d'échange conformationnel, la vitesse de relaxation croisée σ_{IS} tire son origine de l'interaction dipolaire entre les noyaux I et S, modulée par les mouvements de diffusion rotationnelle de la macromolécule. La contribution de la relaxation croisée (dans les conditions de régime initial et d'absence de diffusion de spin) présente l'intérêt non négligeable de varier en fonction de la proximité spatiale des deux noyaux considérés, plus précisément selon un rapport $1/r_{IS}^6$ (r est la distance entre les deux noyaux I et S). Lorsque la distance entre les deux spins est connue et constante, comme c'est le cas par exemple entre un ^1H porté par un ^{13}C ($\sim 1.09 \text{ \AA}$), l'effet NOE hétéronucléaire ^1H - ^{13}C dépend alors principalement du temps de corrélation effectif du dipôle C-H considéré et de l'amplitude de ce mouvement. Il constitue alors un excellent moyen de

quantifier les mouvements internes locaux des différents résidus. L'effet NOE hétéronucléaire est relié aux changements de populations d'un spin S donné qui s'effectuent quand le spin I auquel il est couplé est saturé. Il peut être décrit en utilisant la combinaison des vitesses de relaxation longitudinale $R(C_z)$ et de relaxation croisée $R(H_z \leftrightarrow C_z)$ équation (1) et il est très sensible aux mouvements rapides.

$$NOE = 1 + \frac{\gamma_H}{\gamma_C} \frac{R(H_z \leftrightarrow C_z)}{R(C_z)} \quad (1)$$

où γ_H et γ_C sont les rapports gyromagnétiques respectivement. On peut voir ainsi que dans le cas de l'azote (gamma négatif) le NOE hétéronucléaire est inférieur à 1 mais que dans le cas du carbone (gamma positif) il est supérieur à 1.

La mesure du paramètre NOE 1H - ^{13}C nécessite l'enregistrement de deux expériences : dans l'une, le proton subit une saturation ; dans l'autre, on effectue une saturation hors résonance (qui n'affecte pas le proton) pour compenser d'éventuels effets d'échauffement. Le paramètre NOE 1H - ^{13}C est calculé en appliquant l'équation :

$$NOE = \frac{I}{I_0} \quad (2)$$

où I et I_0 représentent l'intensité des pics de corrélation respectivement avec et sans saturation. Plus les vecteurs H-C sont rigides (ne montrent pas de dynamique interne), plus le paramètre NOE sera proche de 1.

3.1.2 Le paramètre d'ordre S^2

La dynamique interne des résidus d'une macromolécule contribue au phénomène de relaxation des spins qui la composent tout comme d'autres mouvements comme par exemple le mouvement de rotation global de la macromolécule. La mesure des temps de relaxation T_1 , T_2 et de l'effet NOE hétéronucléaire permet d'appréhender les mouvements internes et globaux rapides des macromolécules (dans les gammes de temps de la picoseconde -nanoseconde). Les méthodes de mesure de ces constantes sont traitées avec plus de détail dans le chapitre concernant les mouvements lents. Différents modèles reliant la contribution de la fréquence et de l'étendue du mouvement interne à la fonction d'autocorrélation, qui décrit la fluctuation du champ magnétique local, ont été suggérés (Luginbühl et Wüthrich 2002). L'approche la plus utilisée est celle dite du

« modèle libre » "*model-free*" initialement décrite par Lipari et Szabo (Lipari et Szabo 1982) et généralisée par l'équipe de Bax (Clore et al. 1990). Dans ce modèle (Figure 7) la nature de la dynamique interne n'est pas prise en considération et celle-ci est supposée plus rapide que la vitesse de diffusion rotationnelle de globale de la macromolécule, directement reliée au temps de corrélation global, ce qui implique que le mouvement global et la dynamique interne sont découplés. Le mouvement interne est décrit simplement par le temps de corrélation τ_f et le paramètre d'ordre S_f^2 . La fonction d'autocorrélation pour un type de mouvement interne concernant un résidu est donnée par l'équation :

$$g_i(\tau) = S_f^2 + (1 - S_f^2) e^{(-\tau/\tau_f)} \quad (3)$$

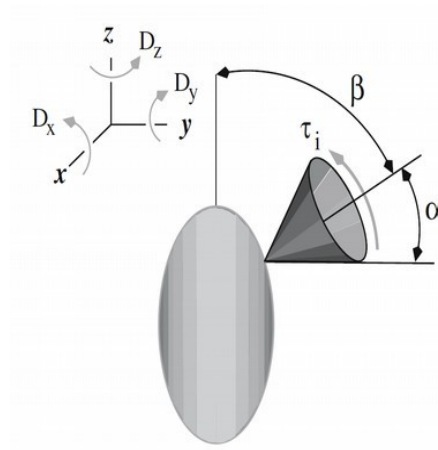


Figure 7 : la diffusion/rotation de la molécule est caractérisée par trois constantes, D_x , D_y et D_z . Pour les molécules à symétrie axiale, comme celle illustrée dans ce schéma, $D_{||} = D_z$ et $D_{\perp} = D_x = D_y$. Le cône représente le mouvement interne du vecteur C-H. La dynamique interne est caractérisée par le demi-angle du cône α et la fréquence de diffusion du vecteur τ_i . L'orientation du vecteur C-H par rapport à l'axe z est représentée par l'angle β .

Si la dynamique interne du vecteur est supposée se produire de façon aléatoire à la surface d'un cône de demi-angle α , le paramètre d'ordre S^2 donné par l'équation :

$$S_f^2 = \left[\frac{1}{2} (1 + \cos \alpha) \cos \alpha \right]^2 \quad (4) \text{ (Luginbühl et Wüthrich 2002)}$$

Plus le mouvement est restreint, plus α tend vers 0 et plus le paramètre d'ordre S^2 se rapproche de 1. Dans le cas contraire, α augmente et S^2 se rapproche de 0.

L'analyse des données avec un modèle libre à un seul régime de dynamique ne tient pas compte des

mouvements de sous-ensembles des résidus qui peuvent correspondre à un régime plus lent a été utilisé en initialement. Les écarts parfois observés entre les données de relaxation mesurées et calculées (Clore et al. 1990) ont effectivement suggéré l'existence d'un mouvement interne plus lent (\sim ns- μ s), associé à un nouveau temps de corrélation τ_s et de paramètre d'ordre S_s^2 . Ce dernier temps de corrélation est néanmoins inférieur au temps de corrélation global τ_c qui constitue donc la limite supérieure des mouvements qui peuvent être détectés par les méthodes de mesure de temps de relaxation. En raison de la présence de ce mouvement dit « lent », la fonction d'autocorrélation est donnée par l'équation :

$$g_i(\tau) = S_f^2 S_s^2 + (1 - S_f^2) e^{(-\tau/\tau_f)} + S_f^2 (1 - S_s^2) e^{(-\tau/\tau_s)} \quad (5)$$

Notons bien que l'accès à des échelles de temps encore plus lentes (microseconde-miliseconde) n'est possible qu'avec les méthodes de dispersion de relaxation (voir chapitre 6)

Cette équation montre que la fonction d'autocorrélation décroît vite selon le terme du mouvement rapide ($1 - S_f^2$) mais subit une diminution lente en fonction du mouvement lent $S_s^2(1 - S_s^2)$. En l'absence de dynamique interne lente ($S_s^2 = 1$), la fonction d'autocorrélation revient au modèle simple à un régime. Dans la pratique, ce sont les paramètres d'ordre relatifs (S_{rel}) qui sont mesurés. Cette mesure consiste en la normalisation de $(2R_2 - R_1)$ sur la plus grande valeur de ce terme pour chaque famille de spins étudiés. Les termes R_2 et R_1 représentent les vitesses de relaxation transversale et longitudinale respectivement. Les méthodes de mesure de ces constantes sont traitées en détail dans le chapitre concernant les mouvements lents.

3.2 Résultats obtenus par relaxation des spins et calculs des paramètres d'ordre S^2

L'étude des mouvements internes se produisant à l'échelle de la pico-nanoseconde a d'abord été effectuée par relaxation T_1 , T_2 des spins et du NOE hétéronucléaire sur des ADN-B non marqués de grande longueur. Les mesures de relaxation du ^{31}P et les NOE hétéronucléaires ^1H - ^{31}P (Klevan, Armitage, et Crothers 1979) ont permis d'estimer que les dynamiques du squelette phosphodiester de l'ADN soit nu soit complexé dans le nucléosome se produisaient à l'échelle de la nanoseconde, plus précisément entre 0.2 à 0.4 ns. Néanmoins, cette interprétation ignorait les contributions liées à la forte anisotropie de déplacement chimique du phosphore. Des études plus approfondies (Bendel et James 1983) des relaxations du ^{13}C et ^{31}P qui tenaient compte de cette anisotropie ont montré sur

des ADN linéaires, circulaires ou super-enroulés que les échelles de temps correspondant aux mouvements des groupements phosphate et des sucres étaient plutôt de l'ordre de 1 à 7ns. Cette gamme de temps avait également été proposée dans d'autres travaux sur des ADN de tailles variables (de 140 à 600 paires de bases), en comparant les relaxations des ^{31}P , ^{13}C et ^1H sous différentes conditions (température, viscosité et champ magnétique) (Hogan et Jardetzky 1980).

Ces travaux ont mis en évidence les dynamiques du squelette phosphodiester et des sucres mais sans pouvoir identifier individuellement les résidus ; En effet avec des ADN de cette taille, toutes les résonances sont complètement superposées. Pour pallier à ces limitations, il est nécessaire d'utiliser des ADN beaucoup plus courts. Le temps de corrélation global (mouvements de diffusion-rotation) de ces oligomères est inférieur à 10 ns.

Les premières mesures de relaxation d'oligomères ont été réalisées dans le début des années 80 (Early et al. 1981a) (Early et al. 1981b) (Feigon et al. 1983). Les valeurs R_1 et R_2 des protons imino ont été mesurées grâce à des expériences dites « d'inversion retour » et « écho de spin Hahn », en faisant l'hypothèse que la relaxation de ces protons est dominée par les interactions dipolaires avec l'azote et que les plus proches protons sont à distance constante du proton d'intérêt. Mais il s'est vite avéré que cette dernière hypothèse n'était pas valable et que les protons voisins contribuaient très vraisemblablement à la relaxation des protons d'intérêt (Lipari et Szabo 1981). Le proton n'est donc pas la sonde idéale pour quantifier la dynamique de l'ADN : tout un réseau de proton existe dans un ADN et les distances entre ces noyaux sont affectées par la dynamique interne de la molécule et les structures locales particulières.

Les calculs des paramètres d'ordre S^2 dans des ADN sont apparus un peu plus tardivement et ont surtout été appliquées aux sucres, et plus particulièrement à l'atome C1'. Dans des petits duplexes d'ADN non marqués mais dont la courte longueur (4 -8 paires de bases) permet une bonne résolution, il a été trouvé que la dynamique est moins importante pour les bases ($S^2 \sim 0.8$) que pour les sucres ($S^2 \sim 0.6$) (Borer et al. 1994). Les mesures des constantes des vitesses de relaxations et des NOE hétéronucléaires des C1' de décarnères (Paquet, Gaudin, et Lancelot 1996), interprétées selon le modèle de Clore (*Extended Model-free*)(Clore et al. 1990), ont conclu à l'existence de deux mouvements internes corrélés sur les sucres, un plus lent ($\tau_s=1.5\text{ns}$, $S_s = 0.74$) que l'autre ($\tau_f=20\text{ps}$, $S_f^2=0.52$ à 0.96). La dynamique des sucres s'est avérée modulée par la séquence, ceux des cytosines étant généralement plus importantes que celles des autres bases (Shajani et Varani 2008) (Duchardt, Nilsson, et Schleucher 2008). Enfin, un travail récent a confirmé l'existence d'un tel effet de séquence (Nikolova et al. 2012). En montrant, toujours avec les mêmes méthodes, que les

pyrimidines des pas CpA, CpG et TpG sont flexibles contrairement à celles des pas ApT et TpT, plutôt rigides. Cette approche a été complétée par une dynamique moléculaire de 10ns réalisée avec le champ de force CHARMM27 qui suggère que les flexibilités des sucres sont reliées aux mouvements des groupements phosphate. Cette conclusion rejoint une étude expérimentale qui montre que les variations des paramètres d'ordre des C1' sont sensibles aux échanges conformationnels du groupement phosphate (Isaacs et Spielmann 2001).

En conclusion, la RMN n'a pour l'instant fournit que des informations partielles sur la dynamique interne des ADN, du moins en ce qui concerne les mouvements rapides. Mais nous allons maintenant exposer comment l'exploitation des données RMN sur l'échange conformationnel des groupements phosphate a permis d'aborder cette question.

4 Les mouvements des groupes phosphate dans l'ADN-B en RMN

4.1 Mise en évidence et quantification de l'équilibre BI ↔ BII

Comme nous l'avons vu dans la section 2.2, il existe deux états conformationnels des groupements phosphate dans l'ADN-B dans les structures cristallographiques, nommés BI et BII, et qui mettent en jeu des angles ϵ et ζ (Figure 3). La conformation BI semblait largement majoritaire et la question était de savoir si ces deux états existaient en solution ou si BII était une curiosité liée aux conditions cristallines.

En RMN, les conformations des angles dièdres θ peuvent être estimées par les constantes de couplages J grâce à l'équation de Karplus (Karplus 1959) $J = A\cos^2\theta + B\cos\theta + C$ (6) où A, B et C sont des constantes empiriques qui dépendent des angles considérés. Dans l'ADN, cette approche est utilisée pour sonder les angles de torsions β ($^3J_{H5'P5'}$, $^3J_{H5''P5'}$ et $^3J_{C4'P5'}$), γ ($^3J_{H4'H5'}$ et $^3J_{H4'P5''}$), ϵ ($^3J_{H3'P3'}$, $^3J_{C4'P3'}$ et $^3J_{C2'P3'}$) du brin phosphodiester et l'angle glycosidique χ ($^3J_{H1'C6/8'}$ et $^3J_{H1'C2/4'}$). Concernant les états BI et BII, en plus du couplage scalaire $^3J_{H3'P3'}$ relié à ϵ , les déplacements chimiques du phosphore 31 (δP) doivent logiquement être sensibles aux équilibres conformationnels des groupements phosphate.

L'équipe de Gorenstein (revue : (David G. Gorenstein 1984) a été la première à montrer que les

variations des déplacements chimiques du ^{31}P et celles des couplages scalaires $^3J_{\text{H3'P3'}}$ étaient corrélées. L'idée s'est ainsi progressivement imposée qu'un équilibre BI \leftrightarrow BII existe en solution et que le déplacement chimique du phosphore reflète le ratio BI/BII.

Mais cette approche ne permettait pas de relier les valeurs des δP à des ratio BI/BII. Pour ce faire, il faudrait idéalement avoir les valeurs des δP pour des groupements phosphate pur-BI et pur-BII. L'approche théorique des δP , basée sur les calculs quantiques d'orbitales moléculaires de petits systèmes sucres-groupements phosphates (David G. Gorenstein 1987), ont proposé les valeurs de -4.6 ppm pour du pur-BI et de -3 ppm pour du pur-BII (référéncé au TriMéthyl-Phosphate). Les δP pur-BI et pur-BII sont donc séparés par 1.6 ppm. En appui de ces résultats, un pas GpT a été expérimentalement caractérisé par un δP de -4.6 ppm et un couplage $^3J_{\text{H3'P3'}}$ de 1.5 Hz, qui selon l'équation de Lankhorst (Lankhorst et al. 1984), $J = 15.3 \cos^2(\epsilon) - 6.1 \cos(\epsilon) + 1.6$ (7), dérivée de la formule de Karplus donne un angle ϵ en trans (autour de 180°), typique d'une conformation BI (Schroeder et al. 1989). Les calculs quantiques (David G. Gorenstein 1987) ont également abouti à mesurer une différence d'énergie entre BI et BII du pas GpT de 1 kcal/mol en faveur de BI, Une étude plus récente (Tian et al. 2009) de mesure de ΔG de plusieurs pas dinucléotidiques a montré qu'un tel écart n'existe pas en faveur de la conformation BII, ce qui implique que la conformation BII ne peut pas être stabilisée à 100% et pourrait expliquer pourquoi il n'a jamais été observé expérimentalement des δP de -3 ppm. Notons tout de même une étude par RMN et simulation de type Monte-Carlo d'un double mésappariement G:A intégré dans une hélice B (Chou, Cheng, et Reid 1992) qui a fait état d'un δP à -2 ppm, interprété comme un pur BII. Mais il est fort probable que cette valeur corresponde à un ensemble de perturbations structurales induites par le double mésappariement et non pas uniquement à l'apparition d'un conformère BII. Beaucoup plus récemment, des calculs quantiques de δP ont été réalisés sur les structures de dinucléotidiques issues d'une dynamique moléculaire du dodécamère de Drew-Dickerson et triées selon l'état du phosphate (Přecechtělová et al. 2013). Ces systèmes sont beaucoup plus sophistiqués que ceux utilisés par l'équipe de Gorenstein, et, de plus, les calculs prennent en compte la solvation (une ou deux couches d'eau). Malheureusement, le champ de force utilisé (Parm99, (Cheatham, Cieplak, et Kollman 1999)) dans les simulations de dynamique moléculaire génère si peu de BII que les auteurs eux-mêmes ne font pas confiance dans les calculs statistiques liés à cet état. De plus, il y a clairement un problème de calibration avec les calculs, les δP calculés n'étant pas du tout dans la même gamme que les δP en RMN. Malgré ces limitations, quelque soit le dinucléotide considéré, la

différence trouvée entre les δP pur-BI et pur-BII est de ~ 1.5 ppm, très proche de la valeur trouvée par Gorenstein. Un point très intéressant soulevé par ces travaux est qu'il pourrait exister un effet de séquence sur les δP pur BI et BII, de l'ordre de 0.4-0.5 ppm. Ainsi, chaque dinucléotide serait caractérisé par des δP pur BI et BII spécifiques, alors que la différence de ~ 1.5 ppm qui sépare les δP pur-BI et pur-BII serait comparable pour tous les pas dinucléotidiques indépendamment de la séquence, donc beaucoup moins marquée par la séquence.

Dans l'état actuel des choses, on ne peut donc pas connaître avec certitude les valeurs des δP des groupements phosphate pur-BI et pur-BII. D'autres stratégies ont donc été mises en place pour pouvoir interpréter les δP mesurés par RMN en termes de populations de BI et BII.

Des simulations de mécanique moléculaire réalisées avec le programme JUMNA (R. Lavery, Zakrzewska, et Sklenar 1995), dont des simulations sous contraintes RMN, ont montré que les distances séquentielles (entre résidu i et résidu $i+1$) impliquant les protons non échangeables des bases (protons H6 et H8) et les protons H2' et H2'' des sucres étaient sensibles aux états BI et BII (Lefebvre et al. 1996). Cette relation entre distances séquentielles et états BI-BII a été confirmée en analysant des structures cristallographiques (Tisné et al. 1998). Remarquons que, dans les deux cas, mécanique moléculaire et structures cristallographiques, les phosphates sont « piégés » soit en BI soit en BII, et qu'il n'y a pas de notion de dynamique. Quoiqu'il en soit, il a été élaboré une première méthode de quantification du taux de BII en solution basée sur les distances expérimentales issues des NOEs en solution ($d(obs)$) et les distances des structures cristallographiques correspondant à des pur-BI ($d(BI)$) et pur-BII ($d(BII)$) (Tisné et al. 1998):

$$\%BI = 100 - \%BII = \frac{d(obs) - d(BII)}{d(BI) - d(BII)} \quad (8)$$

Les corrélations linéaires observées entre les δP et les distances séquentielles H6/8-H6/8, H2'-H6/8 et H2''-H6/8 collectés en RMN sur un oligomère de 14 paires de bases ont ensuite été mises en parallèle avec celles impliquant leurs correspondants extraits de structures cristallographiques résolues à 1 Å ou moins, en substituant juste les δP par les valeurs de $(\epsilon - \zeta)$. En faisant l'hypothèse que le taux de BI décroît linéairement en fonction du terme depuis $(\epsilon - \zeta) = -90^\circ$ (100% de BII) jusqu'à $(\epsilon - \zeta) = +90^\circ$ (0 % de BI) il a été établi la relation empirique

$$BII (\%) = 143 \delta P + 621 \quad (9) \text{ (Heddi et al. 2006).}$$

Trois ans après, le groupe de Hatcher (Tian et al. 2009) a publié une autre méthode de quantification

des ratios BI/BII à partir des δP , fondée sur l'existence d'un échange conformationnel rapide entre les deux états BI et BII, caractérisés par des déplacements chimiques moyens ω_{BI} et ω_{BII} . Le pourcentage de BII est donné par l'équation 4:

$$\frac{\%BII}{100} = \frac{1}{\omega_{BII} - \omega_{BI}} \langle \omega \rangle - \frac{\omega_{BI}}{\omega_{BII} - \omega_{BI}} \quad (10)$$

où $\langle \omega \rangle$ est la moyenne des déplacements chimiques BI et BII observables en RMN. Le problème est ici de déterminer ω_{BI} et ω_{BII} . Pour ce faire, Tian et al ont exploité le comportement des δP de 5 dodécamères en fonction de la température (Figure 8), en faisant l'hypothèse que l'augmentation de la température tend à égaliser les populations BI et BII. Les variations des δP en fonction de la température (Figure 8) ont ainsi été approximées de façon linéaire et extrapolées. Les δP qui augmentaient avec la température ont été utilisés pour prédire ω_{BI} ($\delta P_{BI} = 0.0002031T(K) - 1.345$) et, à l'inverse, les δP décroissant avec la température ont servi à déduire ω_{BII} ($\delta P_{BII} = -0.0003079T(K) + 1.023$)

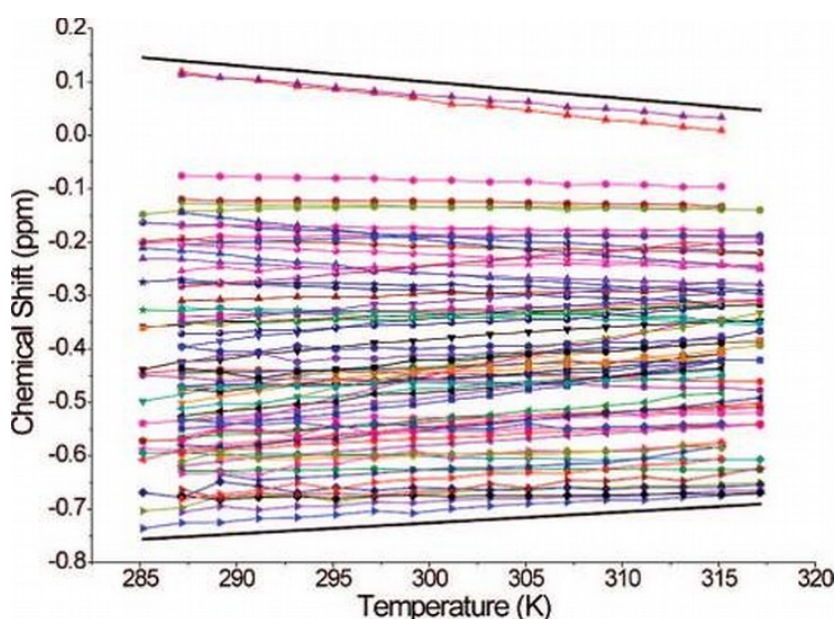


Figure 8: Variations des déplacements chimiques du phosphore 31 (référéncés par rapport à l'acide phosphorique) en fonction de la température pour cinq dodécamères. D'après (Tian et al. 2009)

Bien que les méthodes de Tian et al. et de Heddi et al. soient d'essence très différentes, les résultats sont étonnamment cohérents. Il y a en effet une excellente corrélation (coefficient de corrélation $R=0.98$) entre ces deux interprétations, montrée dans la Figure 2 sur la base de 72 δP que nous

avons collectés sur quatre dodécamères à 30°C. Néanmoins, l'approche de Tian et al donne systématiquement des pourcentages de BII supérieurs à ceux estimés par Heddi et al, notamment pour des phosphates qui peuplent faiblement l'état BII.

Il est très difficile de trouver des critères objectifs pour choisir entre ces deux méthodes. Par exemple, il a été estimé qu'un ratio BI/BII de 1 (50 % BI et 50 % BII) correspondrait à un δP de -4.00 ± 0.02 ppm à partir d'une étude de RMN portant sur une série d'ADN simple brin (Ho et Lam 2004). Mais comme il est montré dans la Figure 9, les deux méthodes d'extrapolation des δP sont en bon accord avec cette valeur.

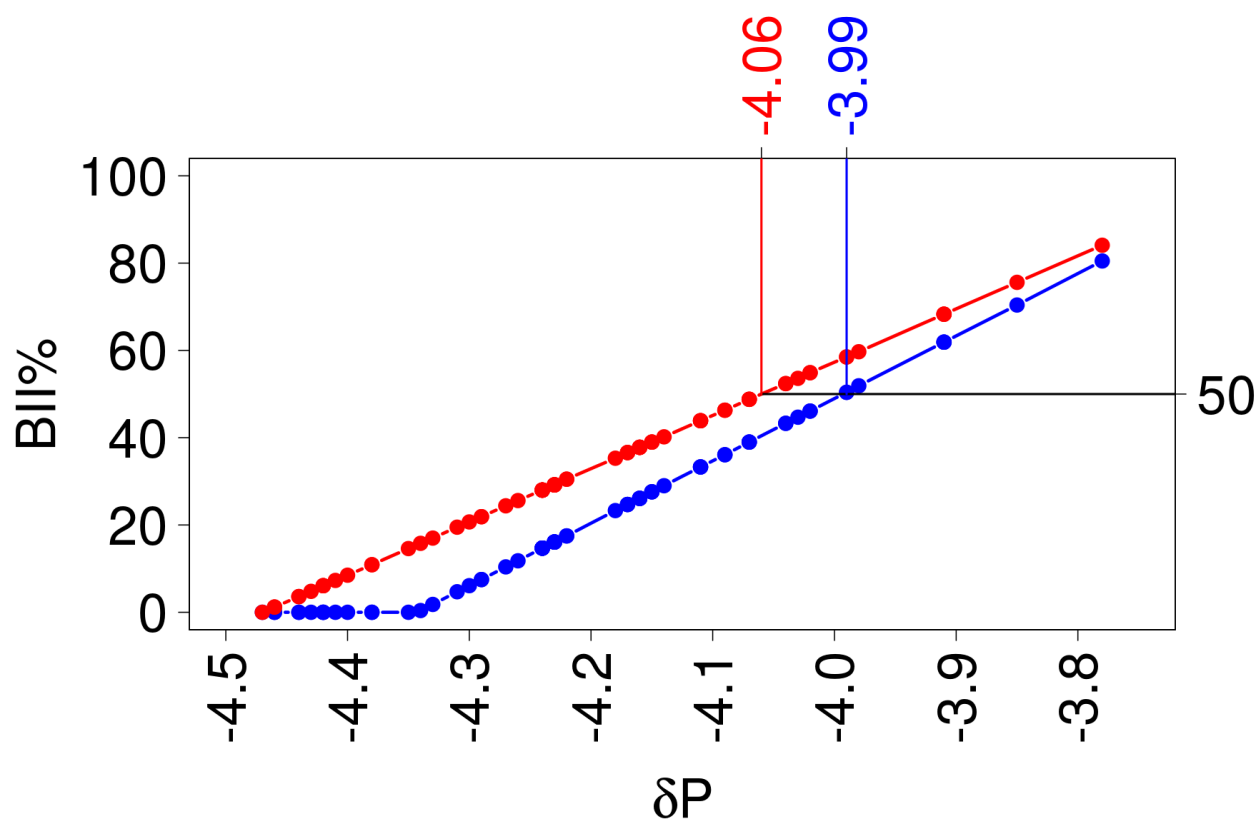


Figure 9 : Pourcentages de BII calculés à partir de 72 δP (ppm) collectés à 30°C au laboratoire, en utilisant soit l'équation de Heddi (Heddi et al. 2006) (bleu), soit la formule de Tian (Tian et al. 2009) (rouge).

4.2 Effet de la séquence de l'ADN sur la dynamique BI \leftrightarrow BII

Il a été constaté assez tôt que les dinucléotides pyrimidine-p-purine semblaient plus favorables au

BII que d'autres types de pas (D. G. Gorenstein 1992). Des travaux RMN ultérieurs ont confirmé que les pas CpG (El antri et al. 1993) (Lefebvre et al. 1995) (Lefebvre et al. 1996) et CpA.TpG (Tisné et al. 1998) (Tisné, Hartmann, et Delepierre 1999) sont particulièrement enclins à adopter la conformation BII. De plus, l'équilibre BI \leftrightarrow BII dans ces dinucléotides est influencé par leur environnement tétranucléotidique (Lefebvre et al. 1996) (Tisné et al. 1998) (Tisné, Hartmann, et Delepierre 1999). Ces résultats ont été étayés par des simulations et l'analyse des structures cristallographiques à moins de 2 Å de résolution (Bertrand et al. 1998).

Mais l'effet de la séquence sur l'équilibre BI \leftrightarrow BII a été généralisé et plus complètement formalisé grâce à la constitution d'une base de données contenant 241 δP qui a permis de montrer que les 16 pas dinucléotidiques constituant l'ADN-B étaient associés à des δP caractéristiques (Figure 10 et Tableau 2) (Heddi et al. 2010). Cette approche a aussi confirmé que le comportement des phosphates des pas CpG, CpA et TpG était modulé par leur environnement, exprimé en terme de pyrimidine (Y) et purine (R). Nous verrons dans cette thèse que la base de données initiale a depuis été enrichie par 72 autres points, pour aboutir aux valeurs présentées dans la Figure 10 et le Tableau 2.

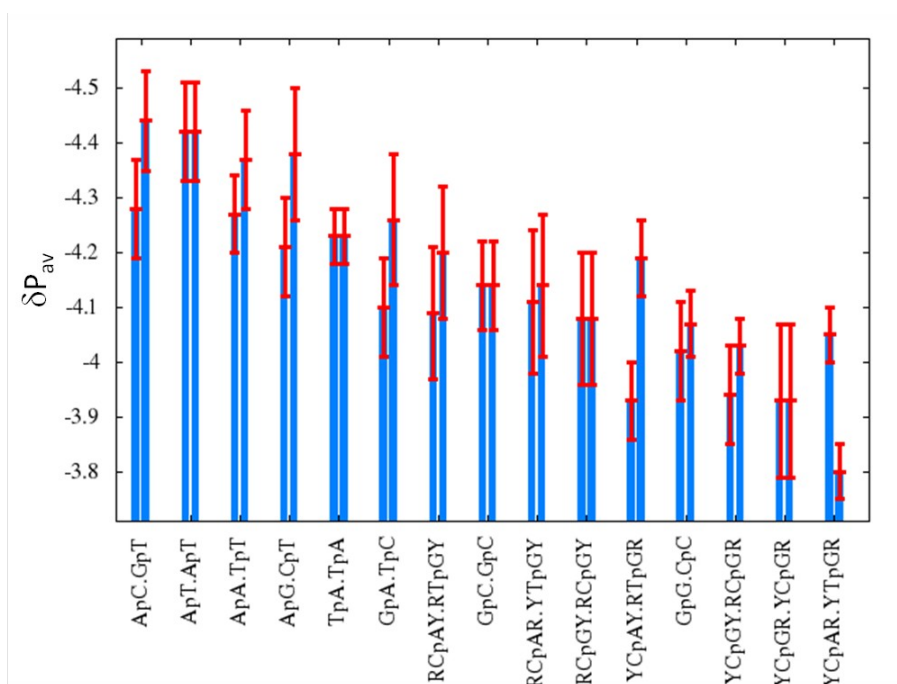


Figure 10 : déplacements chimiques du phosphore (référéncé par rapport au TMP) en fonction de la séquence dinucléotidique et tétranucléotidique pour CpG, CpA et TpG, pyrimidine (Y) et purine (R). Les barres bleues juxtaposées représente les δP des groupements phosphates face à face dans l'ordre de l'annotation.

En extrapolant les δP selon l'approche décrite précédemment (Heddi et al. 2006), il est possible d'exprimer cet effet de séquence en terme de pourcentage de BII, qui reflète plus explicitement la propension à adopter la conformation BII (Tableau 3).

	Step	n	δP_{mov}	$DS_{\delta P}$
Dinucléotides	CpA	31	-3.99	0.16
	GpG	12	-4.02	0.09
	CpG	35	-4.02	0.12
	CpC	12	-4.07	0.06
	GpA	26	-4.10	0.09
	TpG	31	-4.14	0.10
	GpC	34	-4.14	0.08
	ApG	25	-4.21	0.09
	TpA	20	-4.23	0.05
	TpC	26	-4.26	0.12
	ApA	20	-4.27	0.07
	ApC	28	-4.28	0.09
	ApT	22	-4.42	0.09
	GpT	29	-4.44	0.09
	CpT	24	-4.38	0.12
	TpT	20	-4.37	0.09
Effet de l'environnement tétramérique	YCpGR	5	-3.93	0.14
	YCpGY	6	-3.94	0.09
	RCpGR	6	-4.03	0.05
	RCpGY	18	-4.08	0.12
	YCpAR	8	-3.80	0.05
	YCpAY	6	-3.93	0.07
	RCpAY	9	-4.09	0.12
	RCpAR	8	-4.11	0.13
	YTpGR	8	-4.05	0.05
	YTpGY	8	-4.14	0.13
	RTpGR	6	-4.19	0.07
	RTpGY	9	-4.20	0.12

Tableau 2 : Ce tableau donne l'occurrence N pour chacun des 16 dinucléotides complémentaires dans la base de donnée, le déplacement chimique du phosphore moyen correspondant (δP_{mov}), et la déviation standard associée ($DS_{\delta P}$).

Nous avons vu que l'analyse des structures cristallographiques avait montré que les états BI.BI, BI.BII et BII.BII qui peuvent exister au sein des dinucléotides complémentaires étaient associés au déplacement des paires de bases (*X-disp*) mais aussi au *Roll* et au *Twist* (Figure 5 dans la section 1.2). C'est la raison pour laquelle a été conçue l'échelle TRX (pour *Twist* **R**oll et *X-disp* (Heddi et al. 2010) qui attribue à chaque dinucléotidique complémentaire un score qui est égal à la moyenne des

pourcentages des états BII des phosphates face à face (Tableau 3). Un dinucléotidique complémentaire avec un score TRX très faible signifie que les phosphates face à face sont pratiquement piégés en BI. A l'inverse, un score élevé témoigne de phosphates face à face qui partagent chacun leur temps entre BI et BII. L'idée est qu'un dinucléotidique complémentaire riche en BII peut explorer un large paysage conformationnel, couvrant les caractéristiques associées aux trois états BI.BI, BI.BII et BII.BII. Ce type de dinucléotidique sera donc considéré comme plus malléable qu'un dinucléotidique qui est majoritairement en BI.BI.

	%BII		%BII	score TRX
YCpGR	59	YCpGR	59	59 (19)
RCpGR	44	YCpGY	57	51 (14)
RCpGY	38	RCpGY	38	38 (17)
<i>CpG</i>	45	<i>CpG</i>	45	45 (17)
GpG	46	CpC	38	42 (13)
YCpAR	77	YTpGR	42	59 (8)
YCpAY	59	RTpGR	21	40 (11)
RCpAR	33	YTpGY	28	30 (18)
RCpAY	36	RTpGY	19	27 (17)
<i>CpA</i>	50	<i>TpG</i>	29	39 (19)
GpC	30	GpC	30	30 (11)
GpA	34	TpC	12	23 (15)
TpA	15	TpA	15	14 (7)
ApG	19	CpT	0	10 (15)
ApA	10	TpT	0	5 (12)
ApC	8	GpT	0	4 (13)
ApT	0	ApT	0	0 (13)

Tableau 3 : . Ce tableau donne le pourcentage de BII (%BII) de chacun des 16 dinucléotides ainsi que les déviations standards associées, extrapolés des δP_{moy} et $DS_{\delta P}$ du Tableau 2 en utilisant l'équation $BII(\%) = 143 \delta P + 621$ (Heddi et al. 2006). Le score TRX est la moyenne des %BII des phosphate face à face.

L'ensemble du travail sur les δP a pu apporter la première preuve expérimentale qu'en solution la dynamique de l'ADN, au moins celle des phosphates, dépend étroitement du contexte de séquence. Mais l'établissement de l'échelle TRX veut aussi dire que la dynamique du phosphate peut être prédite à la simple lecture des dinucléotides composant la séquence d'intérêt. Enfin, si les couplages entre état conformationnels des phosphates et paramètres hélicoïdaux mis en lumière sur les structures cristallographiques existent aussi en solution, il devient possible d'envisager de prédire la variabilité de la double hélice B au niveau dinucléotidique le long de n'importe quelle séquence.

Une autre méthode qui, idéalement, devrait pouvoir prédire le comportement dynamique des ADN est la simulation numérique, et en particulier la dynamique moléculaire. Nous allons maintenant présenter cette approche, et voir comment les résultats expérimentaux obtenus sur la dynamique des des groupements phosphate a été et continue d'être exploitée pour tester et améliorer les champs de force.

5 Simulations de la dynamique des ADN à l'échelle de piconanoseconde

Cinq ans après la publication de la première dynamique moléculaire d'une protéine en 1977 par l'équipe de Karplus (McCammon, Gelin, et Karplus 1977), le même groupe réalisa la première simulation d'un petit ADN de séquence palindromique d(CpGpCpGpCpG)₂ (Tidor et al. 1983). La même année, Michael Levitt (Levitt 1983) publia une simulation du dodécamère de Drew-Dickerson.

Depuis, l'approche théorique de la dynamique des ADN a beaucoup évolué et intéresse de très près la communauté scientifique. Une simple recherche dans le site *Web Of Science* (<http://apps.webofknowledge.com>) avec les mots clés '*DNA SIMULATION*' révèle plus de 1000 publications scientifiques contiennent ces deux mots dans leurs titres entre 2005 et 2015. L'intérêt pour cette méthode vient du fait qu'elle peut apporter des informations au niveau atomique sur la structure et la flexibilité des macromolécules, dont l'ADN, qui, pour l'instant, échappent aux approches expérimentales, comme la RMN ou la cristallographie.

De plus, les progrès technologiques en matière de puissance de calculateurs et les améliorations réalisées sur les logiciels, bien adaptés à la parallélisation sur plusieurs processeurs et maintenant plus conviviaux, contribuent à la notoriété des simulations de dynamique moléculaire. A l'heure actuelle, les simulations sont plus réalistes dans le sens où elles sont toutes faites en présence d'eau et d'ions et que leur durée peut atteindre quelques dizaines de microseconde (Galindo-Murillo, Roe, et Cheatham III 2015). Dans cette partie du manuscrit, nous allons discuter de l'avancée des méthodes de simulation, ainsi que leurs limitations. Mais avant d'aborder ce sujet, nous ferons une digression sur les modélisations réalisées sous la contrainte de données RMN.

5.1 Le cas particulier des affinements des structures d'ADN à partir de données RMN

De nombreux algorithmes (Kuntz et al. 1982) (Havel et Wüthrich 1984) ont été mis au point dès les années 80 pour déterminer les structures d'ADN à partir d'un jeu de données RMN, qui, pendant longtemps, a consisté essentiellement en distances. Un exemple de ces premières approches est la structure du dodécamère de Drew-Dickerson obtenue à partir de distances RMN (Nerdal, Hare, et Reid 1989). La méthode d'affinement était basée sur un algorithme géométrique minimisant l'écart entre distances du modèle de départ et distances expérimentales (Hare et Reid 1986). Cependant les structures résultantes n'étaient pas conformes à la double hélice B, comme par exemple des valeurs de *Twist* extrêmement faibles ($\sim 6^\circ$) pour certains pas CpG.

Une approche intéressante, plus tardive, a consisté à utiliser l'algorithme JUMNA (JUNCTION Minimization of Nucleic Acids) (Richard Lavery 1988). Cet algorithme de mécanique moléculaire minimise l'énergie d'une conformation de départ en utilisant des coordonnées internes (uniquement des angles) et non pas des coordonnées atomiques, ce qui permet d'accélérer considérablement les calculs. L'échantillonnage est assuré en multipliant les points de départ. Le solvant est traité de manière implicite, et, de plus, les charges sur les groupes phosphate sont diminuées. Cette approche a permis de mettre en lumière pour la première fois que les distances expérimentales étaient beaucoup mieux reproduites si, pour certains pas comme les pas CpG et CpA, on considérait deux modèles qui différaient par la conformation des brins, BI dans un cas, et BII dans l'autre (Mauffret et al. 1992) (Lefebvre et al. 1995) (Lefebvre et al. 1996) (Tisné et al. 1998) (Tisné, Hartmann, et Delepierre 1999).

Le développement de la dynamique moléculaire a après quoi gagné le monde des structures d'ADN. Mais il est connu que les distances collectées sur les ADN, à cause de superpositions fréquentes, ne sont pas également réparties le long de la séquence. En conséquence, certains pas dinucléotidiques sont bien contraints alors que d'autres sont livrés au champ de force. Un moyen de pallier à cette inhomogénéité consiste à dériver des distances des δP (Heddi et al. 2008) (Abi-Ghanem et al. 2009). Cette stratégie permet d'améliorer nettement la précision des structures obtenues. Mais, dans ce domaine, une avancée décisive a été l'enrichissement des données RMN par les Couplages Résiduels Dipolaires (RDC). En 2000, le groupe de Bax (Tjandra et al. 2000) a utilisé un nombre record de contraintes RMN (162 NOE, 44 angles de torsion issus des couplages scalaires et 208

RDCs), d'ailleurs jamais dépassé jusqu'à nos jours, dans une courte dynamique moléculaire du dodécamère de Drew-Dickerson. Des contraintes « artificielles » ont été rajoutées sur les appariements des paires de bases et sur les angles de brin, dont ζ , contraint en g-, c'est à dire en BI. L'affinement se fait en plusieurs étapes, en partant de structures soit « *random coil* » soit en conformation A ou B, jusqu'à des structures contraintes par l'ensemble des données RMN dans une dynamique moléculaire de 25ps. Aucune précision (traitement du solvant, champ de force, etc.) n'est donnée dans l'article qui rapporte cette dynamique, à part le fait qu'elle a été réalisée avec le logiciel XPLOR. Les auteurs montrent que l'introduction des RDC est fondamentale pour accroître la précision des structures. De fait, les structures de plus basse énergie issues des différents points de départ sont très semblables entre elles (RMSD de 0.1Å) et également proches des structures cristallographiques du dodécamère de Drew-Dickerson (RMSD de 1Å ou moins). Les sucres sont tous en conformation Sud, et les phosphates en conformation BI. D'une manière générale, l'ADN résultant présente une structure très régulière, sans déviations notables par rapport à un ADN-B canonique. Ce type de protocole a servi de référence pour d'autres études, comme celle de la courbure d'un tract de A contenu dans un dodécamère (MacDonald et al. 2001) ou de la structure d'un oligomère de 17 paires de bases (Olivier Mauffret, Tevanian, et Fermandjian 2002).

Néanmoins, un autre travail effectué sur le dodécamère de Drew-Dickerson soulève quelques questions au sujet de ce protocole (Schwieters et Clore 2007). Dans cette étude les modélisations prennent en compte les NOE, les couplages scalaires et les RDC publiés par le groupe de Bax (Tjandra et al. 2000) mais, cette fois, sans aucune contrainte sur les transitions BI \leftrightarrow BII. Les structures résultantes sont ensuite confrontées aux données de diffusion des rayons X aux grands angles précédemment obtenus (Zuo et Tiede 2005) et à des paramètres d'ordre S^2 expérimentaux (Boisbouvier et al. 2003) Les auteurs montrent qu'une seule structure ne suffit pas à reproduire ces données et suggèrent ainsi qu'il est fondamental de considérer la dynamique de l'ADN, c'est à dire un ensemble de structures représentatives. Il faut aussi souligner qu'une excellente corrélation est trouvée entre les pourcentages de BII issue de la dynamique moléculaire sous contraintes RMN et les prédictions TRX décrites au paragraphe 4.2 (Figure 11).

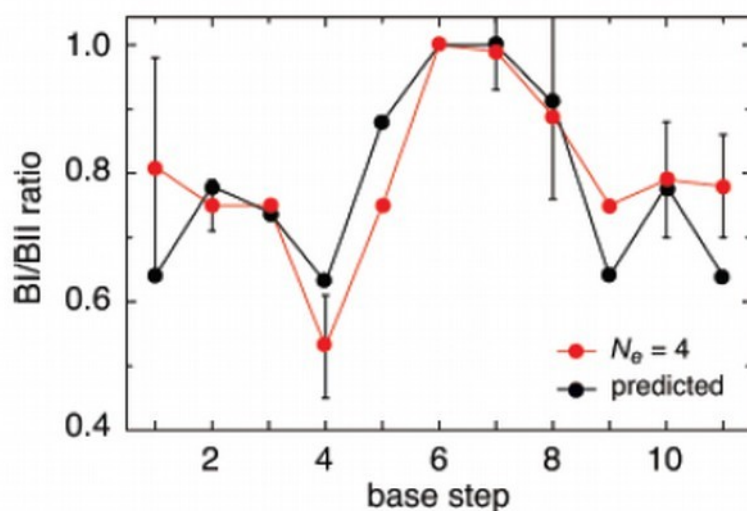


Figure 11 : Ratio BI/BII en fonction de la séquence du dodécamère de Drew-Dickerson. En noir les résultats des simulations et en rouge les ratio issus des prédictions TRX (Heddi et al. 2006). D'après (Schwieters et Clore 2007).

En conclusion, bien qu'une nette amélioration des modèles d'ADN issus des données RMN a été réalisée, en particulier avec l'accroissement du nombre de contraintes, il semble que les protocoles d'affinements ne soient pas toujours aussi fiables qu'on pourrait l'espérer. Sous-jacent à ce problème, se pose la question des champs de force, que nous allons aborder.

5.2 Dynamiques moléculaires et champs de force empiriques

Un champ de force est un jeu de paramètres et d'équations décrivant les interactions entre atomes liés et non liés dans un système moléculaire. Les interactions entre atomes liés correspondent à des énergies de déformation des liaisons, des angles de valence, des angles de torsion et des angles dièdres impropres. Les interactions entre atomes non-liés sont représentées par les énergies de van der Waals et électrostatiques ainsi que par l'énergie des liaisons hydrogène.

Depuis les années 1995, les champs de force traitent les effets électrostatique longue distance correctement, en particulier grâce à l'approche « *Particle Mesh Ewald* » (Darden, York, et Pedersen 1993)). Cette étape est décisive pour les acides nucléiques, qui sont par essence des macromolécules très chargées. Jusqu'à ces dernières années, les principaux représentants de ces champs de force modernes spécifiquement conçus pour pouvoir traiter les ADN en plus des protéines ont été CHARMM27 (Foloppe et MacKerell 2000) (MacKerell, Banavali, et Foloppe 2000) (MacKerell,

Banavali, et Foloppe 2001) pour la famille CHARMM et Parm94 (Cornell et al. 1996) pour la famille AMBER.

Profitant des avancées en matière de logiciels et de puissance des ordinateurs, le consortium ABC (pour « Ascona B-DNA Consortium ») qui s'est constitué en 2001, s'est donné pour tâche de relier la séquence à la malléabilité de l'ADN en étudiant les 136 tétranucléotides possibles en simulant 39 oligomères de 18 paires de bases avec les champs de force Parm. Un premier article (Beveridge et al. 2004) utilisant Parm94 suggérait que les pas YpR, RpR et RpY pourraient être classés selon les profils de paramètres hélicoïdaux HTP/LTP (voir section 2.2) (Yanagi, Privé, et Dickerson 1991). Mais il s'avérait aussi que Parm94 ne donnait pas satisfaction au niveau i) des *Twists*, beaucoup trop faibles, ii) des transitions BI ↔ BII, beaucoup trop rares et iii) des angles α/γ qui passaient irréversiblement en t/g- (Beveridge et al. 2004) (Pérez et al. 2007), une conformation non canonique jamais rencontrée dans les structures cristallographiques (Djuranovic et Hartmann 2003) (Varnai 2002). Parm98 et Parm 99 (Cheatham, Cieplak, et Kollman 1999), qui ont succédé à Parm94, ont donné de meilleurs résultats sur les *Twists*, mais n'ont pas réglé le problème sur les angles du brins (Varnai 2002) (Dixit et al. 2005) (Pérez et al. 2007). La réalisation de dynamiques longues (200 ns) a montré que les transitions t/g- des angles α/γ sont particulièrement problématiques puisqu'elles se propagent tout le long du squelette de l'ADN et finissent par tordre complètement l'ADN (Pérez et al. 2007).

Ce biais a été rectifié dans Parmbsc0 (Pérez et al. 2007). Le groupe ABC réalisa alors de nouvelles dynamiques d'1 μ s sur les 39 oligomères contenant 136 tétranucléotides constitutifs de l'ADN avec Parmbsc0 (Pasi et al. 2014). Un des résultats de ce travail est que la durée des simulations permet d'augmenter les pourcentages de BII (Figure 12). Cependant, ces pourcentages ne respectent pas l'effet de séquence établi expérimentalement par RMN (Heddi et al. 2010). Ce biais est particulièrement visible sur les pas CpG, CpA et TpG, bien moins favorables à BII que d'autres dinucléotides avec Parmbsc0 (Figure 12), alors que les δP indiquent qu'ils sont parmi les pas les plus favorables au BII (Tableaux 2 et 3). Il a été montré par ailleurs que seules l'introduction de contraintes de distances RMN pouvait améliorer de l'effet de séquence sur les transitions BI ↔ BII (Heddi et al. 2008).

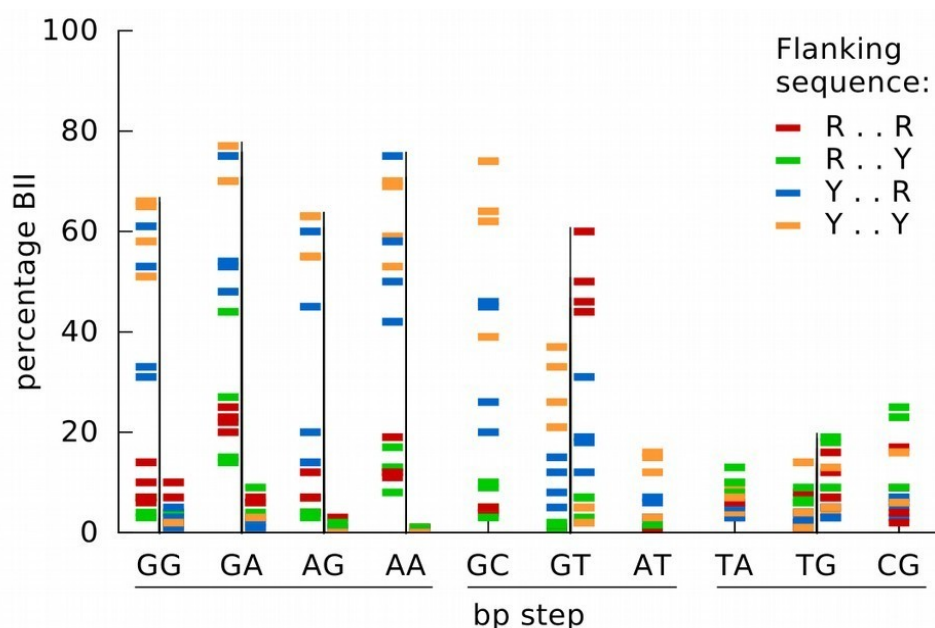


Figure 12 : Moyennes des pourcentages de BII des 10 dinucléotides possibles sur des trajectoires de dynamique moléculaire d'une microseconde. Les lignes verticales noires séparent les pourcentages de BII des pas face à face (à gauche le pourcentage de BII du pas mentionné en abscisse et à droite le pas complémentaire) sauf pour les dinucléotides palindromiques GC, AT, TA et CG, dont les moyennes sont confondues. Les voisins des pas dinucléotidiques sont représentés par les 4 différentes couleurs indiquées dans la légende à droite (R pour purine et Y pour pyrimidine). D'après (Pasi et al. 2014).

Il a donc fallu apporter de nouvelles améliorations au champ de force Parmbsc0 en modifiant la paramétrisation des angles ϵ et ζ (Parmbsc0 _{$\epsilon\zeta$ OLI}) (Zgarbová et al. 2013). Les auteurs de cette publication ont confronté les résultats de leurs dynamiques moléculaires du dodécamère de Drew-Dickerson aux résultats expérimentaux (Figure 13) et montrent que Parmbsc0 _{$\epsilon\zeta$ OLI} améliore sensiblement la représentation de la dynamique du phosphate.

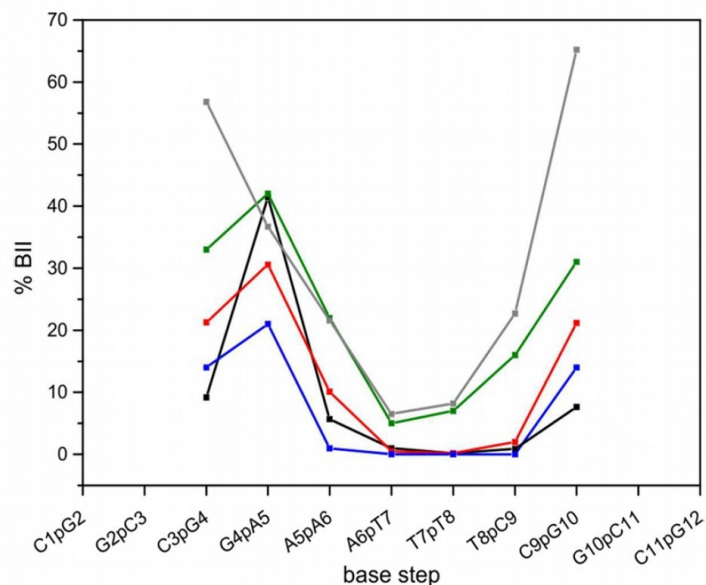


Figure 13: Pourcentages de BII le long du dodécamère de Drew-Dickerson : simulations (1 μ s) avec Parmbsc0 (noir) ou Parmbsc0 _{ϵ GOLI} (rouge), données expérimentales en gris (Tian et al. 2009), en vert (Schwieters et Clore 2007) et en bleu (Heddi et al. 2006). D'après (Zgarbová et al. 2013).

De son côté, CHARMM27 sous-estimait les conformations BII encore plus que Parmbsc0 (Heddi et al. 2008). CHARMM36 (Hart et al. 2012) a représenté une grande avancée dans cette famille de champ de force, comme en témoigne la Figure 14.

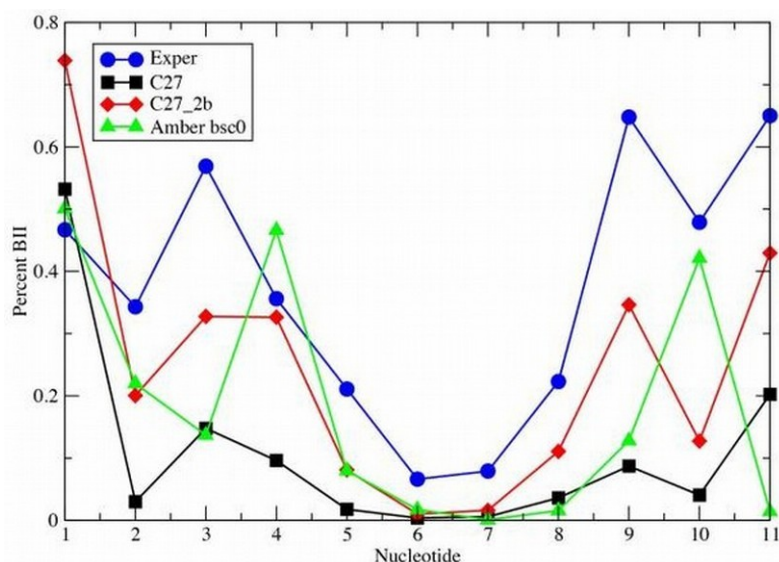


Figure 14: Pourcentage de BII le long du dodécamère de Drew-Dickerson : simulations (100ns) avec Parmbsc0 (vert), CHARMM27 (noir) et CHARMM36 (rouge) ; données RMN en bleu (Tian et al. 2009). D'après (Hart et al. 2012).

Avec Parmbsc0_{εCOLI} et CHARMM27, on peut noter que les pourcentages de BII simulés, bien que nettement supérieurs à leurs parents, restent inférieurs à leurs équivalents expérimentaux (Figures 13 et 14). Néanmoins, ces résultats sont très encourageants et vont dans le sens d'une amélioration dans la représentation de l'effet de séquence sur la dynamique des brins de l'ADN. Mais il est clair que ces champs de force doivent être testés sur d'autres séquences que celle du dodécamère de Drew-Dickerson avant de pouvoir statuer plus fermement sur leur capacité à reproduire les données expérimentales.

Deux questions restent ouvertes, qui semblent être liées. La première concerne la durée des simulations nécessaire pour assurer la convergence des dynamiques. La convergence a été étudiée sur de très longues dynamiques d'environ 90 μ s d'un oligomère de 18 paires de bases simulé avec les champs de force Parmbsc0 et CHARMM36 (Galindo-Murillo, Roe, et Cheatham III 2015) . Pour les 10 paires de bases internes, la convergence, suivie par des mesures de RMSD et de fluctuations atomiques, est déjà excellente dans les premières 300ns. Par contre, la convergence des 4 premières et 4 dernières paires de bases n'est toujours pas parfaite au bout de 90 μ s. Ce phénomène semble lié à une instabilité des plateaux de bases terminaux, qui peuvent se propager aux voisins. Ces bases ouvertes sont soumises à des mouvements variés, de grande amplitude et durant chacun quelques centaines de nanosecondes. Il se crée ainsi des contacts avec des phosphates internes ou les sillons de l'ADN (exemples dans la Figure 15).

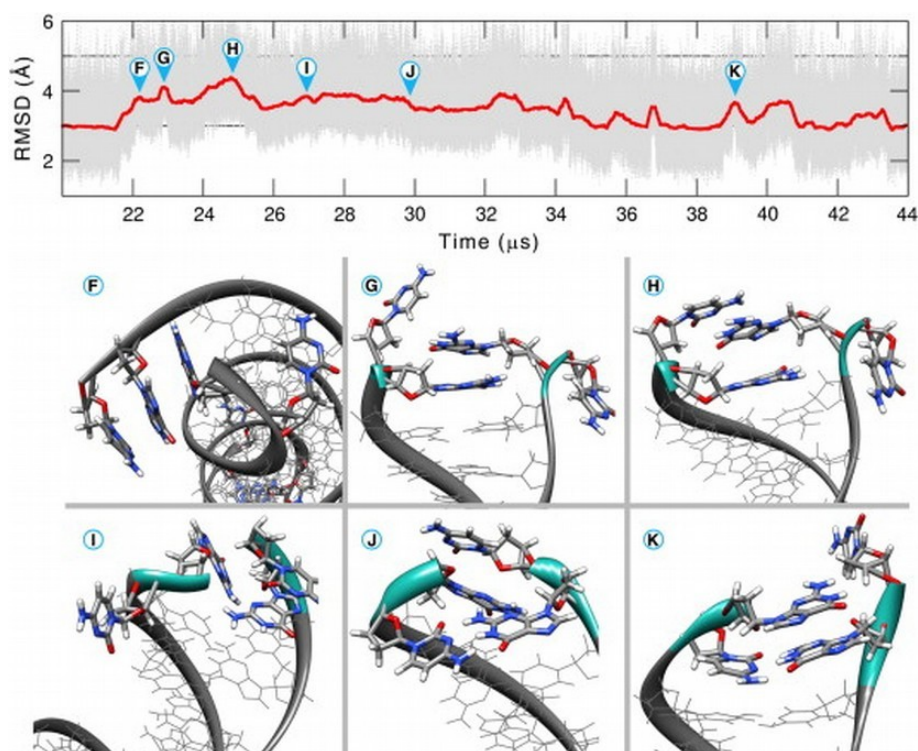


Figure 15 : Exemples de quelques mouvements de bases terminales au cours d'une dynamique moléculaire avec le champ de force CHARMM36. Le graphique du haut représente l'évolution du RMSD (Root Mean square deviation) pour 20 μ s de dynamique moléculaire, calculés et moyennés sur des fractions de temps de 50 ps (en gris) ou de 250ns (en rouge). Les zones du RMSD annotées F, G, H, I, J et K correspondent aux conformations des bases terminales illustrées dans les volets du bas de la figure. D'après (Galindo-Murillo, Roe, et Cheatham III 2015)

Une autre étude de dynamique moléculaire avec Parmbsc0 (Zgarbová et al. 2014) répertorie également ce genre de mouvements. Mais les auteurs suggèrent que ces phénomènes ne sont pas réalistes et qu'ils sont dus à des artefacts du champ de force, plus particulièrement une mauvaise paramétrisation de l'angle glycosidique χ .

En résumé, les approches de dynamique moléculaire, à travers les améliorations des champs de force, commencent à pouvoir fournir des informations précieuses sur la dynamique de l'ADN. Il reste malgré tout à mieux s'assurer de leur réalisme, tant au niveau de l'effet de séquence qu'au sujet du comportement des extrémités des oligomères.

6 L'étude des mouvements lents dans l'ADN par RMN en solution

A l'heure actuelle, les mouvements lents de l'ADN sont encore très peu étudiés. Seul le groupe de Al-Hashimi a abordé cette question en utilisant la RMN, seule technique expérimentale susceptible de fournir ce type d'information à l'échelle atomique. Nous allons décrire en détail leur stratégie, dans la mesure où elle est très récente et que nous l'avons reprise au cours de cette thèse.

6.1 Introduction à la relaxation des spins

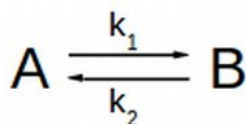
La relaxation des spins dans une molécule se produit sous l'effet d'un champ magnétique externe et dépend de la fréquence à laquelle les champs magnétiques localement ressentis par les spins fluctuent en magnitude et en direction. La fluctuation des champs magnétiques localement induits dépend du mouvement global de la molécule ainsi que des mouvements, plus rapides, associés à sa dynamique interne. Les mesures des paramètres de relaxation constituent donc des données très puissantes pour sonder la dynamique interne des macromolécules, protéines et acides nucléiques.

De façon très générale, sans rentrer dans le détail des mécanismes impliqués dans la relaxation des spins, la vitesse de retour à l'état d'équilibre d'un système de spins $1/2$, après une impulsion radiofréquence, peut s'exprimer sous forme de deux constantes de vitesse de relaxation : la vitesse de relaxation longitudinale R_1 (retour à l'aimantation d'équilibre dans la direction parallèle à la direction du champ magnétique statique) et la vitesse de relaxation transversale R_2 (retour à l'aimantation d'équilibre dans le plan perpendiculaire au champ magnétique statique). L'exposé des mécanismes physiques impliqués dans le processus de relaxation des spins est décrit dans de nombreuses revues dont (Solomon 1955) (Luginbühl et Wüthrich 2002) auxquelles nous renvoyons, de tels exposés dépassant largement le cadre de cette présentation. On peut toutefois mentionner les deux mécanismes physiques les plus importants dans les macromolécules biologiques: il s'agit des interactions dipôle-dipôle (DD) qui sont significatives pour des spins proches dans l'espace et l'anisotropie de déplacement chimique associée aux modulations de l'orientation des noyaux par rapport au champ magnétique statique. En outre, la vitesse de relaxation transversale R_2 peut être affectée par les processus d'échange conformationnel. C'est ce dernier aspect exclusivement qui nous intéresse dans ce chapitre

6.2 L'échange conformationnel et la relaxation des spins

L'échange conformationnel entre deux ou plusieurs environnements entraîne des effets importants sur les vitesses de relaxation et modulent la forme et la position des résonances des noyaux en échange. De ce fait, sous certaines conditions, que je vais exposer dans ce manuscrit, il est possible d'obtenir par la mesure de ces effets, des informations sur les paramètres thermodynamiques, cinétiques et structuraux associées aux états en échange conformationnel.

Lorsqu'un spin s'échange entre deux environnements -le cas le plus simple- son déplacement chimique et sa vitesse de relaxation peuvent différer dans chaque environnement.



La cinétique de l'échange est définie par les constantes de vitesse d'échange, qui sont les constantes de vitesse d'échange de l'état A vers l'état B (k_1) et la réaction inverse (k_2).

Les fractions du système dans chaque conformation est définie par p_A et p_B :

$$p_A = \frac{k_2}{k_1 + k_2} \quad p_B = \frac{k_1}{k_1 + k_2} \quad (11)$$

Dans les études RMN des échanges conformationnels, on considère la vitesse d'échange $k_{ex} = k_1 + k_2$ (12) et la différence des fréquences de précession des deux états en échange ω_A et ω_B ; $\Delta\omega = \omega_A - \omega_B$ (13).

Dans la suite de ce chapitre je vais expliciter la nature des relations existant entre les mesures RMN et ces différents paramètres (populations et fréquences des deux états, vitesses de conversion entre les deux états).

6.2.1 Théorie générale de l'échange conformationnel ("*chemical exchange*") dans le cas d'un système à deux états

L'échange conformationnel à deux états est donné par l'équation suivante (Palmer, Kroenke, et Patrick Loria 2001):

$$\frac{d}{dt} \begin{bmatrix} [A] \\ [B] \end{bmatrix} = \begin{bmatrix} -p_B k_{ex} & p_A k_{ex} \\ p_B k_{ex} & -p_A k_{ex} \end{bmatrix} \begin{bmatrix} [A](t) \\ [B](t) \end{bmatrix} \quad (14)$$

Pour l'aimantation transverse sujette à la libre précession, l'Hamiltonien de Zeeman (\hat{H}) décrit l'interaction des spins avec le champ magnétique externe B_0 : $\hat{H} = \gamma \hat{S}_z B_0$ (15) où γ est le rapport gyromagnétique et \hat{S}_z est le moment angulaire de spin . L'évolution de l'aimantation est décrite pas l'équation modifiée de McConnell (McConnell 1958) :

$$\frac{d}{dt} \begin{bmatrix} M_A^+(t) \\ M_B^+(t) \end{bmatrix} = \begin{bmatrix} -i\omega_A - R_{2A}^0 - p_B k_{ex} & p_A k_{ex} \\ p_B k_{ex} & -i\omega_B - R_{2B}^0 - p_A k_{ex} \end{bmatrix} \begin{bmatrix} M_A^+(t) \\ M_B^+(t) \end{bmatrix} \quad (16)$$

La solution de cette équation différentielle est :

$$\begin{bmatrix} M_A^+(t) \\ M_B^+(t) \end{bmatrix} = \begin{bmatrix} a_{AA}(t) & a_{AB}(t) \\ a_{BA}(t) & a_{BB}(t) \end{bmatrix} \begin{bmatrix} M_A^+(0) \\ M_B^+(0) \end{bmatrix} \quad (17)$$

où :

$$\begin{aligned} a_{AA}(t) &= \frac{1}{2} \left[\left(1 - \frac{-i\Delta\omega + R_{2A}^0 - R_{2B}^0 + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) e^{(-\lambda_- t)} + \left(1 + \frac{-i\Delta\omega + R_{2A}^0 - R_{2B}^0 + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) e^{(-\lambda_+ t)} \right] \\ a_{BB}(t) &= \frac{1}{2} \left[\left(1 + \frac{-i\Delta\omega + R_{2A}^0 - R_{2B}^0 + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) e^{(-\lambda_- t)} + \left(1 - \frac{-i\Delta\omega + R_{2A}^0 - R_{2B}^0 + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) e^{(-\lambda_+ t)} \right] \\ a_{AB}(t) &= \frac{k_{ex} p_A}{\lambda_+ - \lambda_-} [e^{(-\lambda_- t)} - e^{(-\lambda_+ t)}] \\ a_{BA}(t) &= \frac{k_{ex} p_B}{\lambda_+ - \lambda_-} [e^{(-\lambda_- t)} - e^{(-\lambda_+ t)}] \quad (18) \end{aligned}$$

et

$$\lambda_{\pm} = \frac{1}{2} \{ -i\omega_A - i\omega_B + R_{2A}^0 + R_{2B}^0 + k_{ex} \pm [(-i\Delta\omega + R_{2A}^0 - R_{2B}^0 + k_{ex}(p_B - p_A))^2 + 4p_A p_B k_{ex}^2]^{1/2} \} \quad (19)$$

$M_A^+(t)$ et $M_B^+(t)$ sont les composantes transversales de l'aimantation des deux états A et B respectivement. R_{2A}^0 et R_{2B}^0 sont les constantes de relaxation transversale intrinsèque (sans prendre en compte le phénomène d'échange conformationnel) des deux états A et B. Le spectre RMN résultant est obtenu après transformée de Fourier de $M_A^+(t) + M_B^+(t)$.

Les spectres simulés (Palmer, Kroenke, et Patrick Loria 2001), à partir de cette équation, sont représentés dans la figure 16. Deux cas de figure sont envisagés dans cette simulation : le premier est un échange symétrique où $p_A = p_B = 0.5$, le deuxième est un échange asymétrique où $p_B = p_A/4$. En absence d'échange conformationnel, deux raies bien résolues sont observées avec des résonances ω_A

et ω_B (f et l), et des fréquences de relaxation R_{2A}^0 et R_{2B}^0 . Sous l'effet de l'augmentation de la vitesse d'échange (dans la Figure 16, la vitesse augmente du bas vers le haut) les raies s'élargissent et se décalent jusqu'à leur coalescence quand $k_{ex}=\Delta\omega$ (Figure 16 c et i) . Quand l'échange devient très rapide une seule raie est observée à une fréquence $\omega=p_A\omega_A+p_B\omega_B$ (a et g) et la vitesse de relaxation moyenne R_{2avg} est donné par la relation $R_{2avg}=p_AR_{2A}^0+p_BR_{2B}^0$ (20).

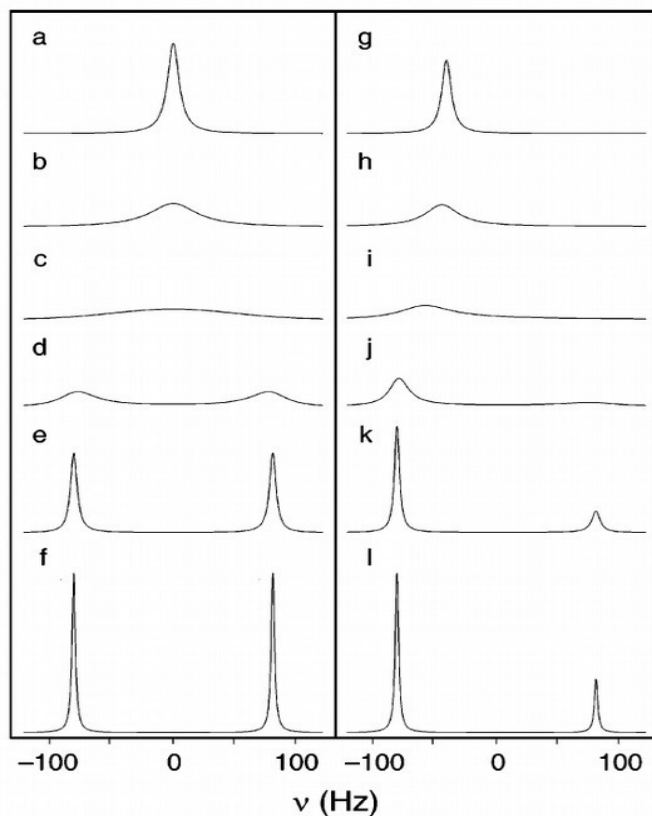


Figure 16: Résonances associées à la présence d'un échange conformationnel symétrique (a-f) $p_A=p_B=0.5$ et asymétrique (g-l) $p_B=0.75$, $p_A=0.25$. Les valeurs de k_{ex} sont : (a,g) 10000, (b,h) 2000, (c,i) 900 (d,j) 200 (e,k) 20 et (f,l) 0Hz. Les spectres sont simulés avec $R_{2A}^0=R_{2B}^0=10\text{Hz}$ et $\Delta\omega=180\text{Hz}$.

6.2.2 L'échange rapide ($k_{ex} \gg \Delta\omega$)

Dans les conditions d'échange rapide, une seule résonance est visible dans les spectres (Figure 16, a g, b et h). Le déplacement chimique est moyenné selon le poids des populations car le temps de vie de chaque état n'est pas assez long pour permettre l'établissement d'un déplacement chimique caractéristique associé à cette conformation. Par conséquent la fréquence observée est $\omega_{obs} =$

$$p_A \omega_A + \omega_B p_B \quad (21).$$

La conséquence visible de ce phénomène est l'élargissement des raies du spectre RMN résultant. Cet élargissement est dû à la contribution positive de l'échange à la relaxation (R_{ex}), où

$$R_{ex} = p_A R_{2A}^0 + p_B R_{2B}^0 + \frac{p_A p_B \Delta \omega^2}{k_{ex}} \quad (22) \text{ (Davis, Perlman, et London 1994)}$$

L'échange augmente la vitesse de relaxation transverse effective (R_{eff}), qui est dans ce cas, la somme de la vitesse de relaxation transversale intrinsèque R_2^0 et de la contribution de l'échange à la relaxation R_{ex} ($R_{eff} = R_2^0 + R_{ex}$)

Comme le montre l'équation (22) le terme R_{ex} est relié à la différence des fréquences entre les deux états $\Delta \omega^2$ et la constante de l'échange. Plus k_{ex} est important dans la limite de l'échange rapide (Figure 16 b et h), plus la contribution de l'échange à la relaxation est notable.

6.2.3 L'échange lent ($k_{ex} \ll \Delta \omega$)

C'est le cas limite opposé, dans lequel la différence des fréquences de précession entre les deux états A et B est plus grande que la vitesse d'échange $k_{ex} = k_1 + k_2$ entre les deux états (Figure 16 e et k). Sous ces conditions, les déplacements chimiques ne sont pas moyennés et deux résonances discrètes correspondant aux fréquences des états A et B sont observées. L'intensité relative de ces résonances est proportionnelle à la fraction du système dans chaque conformation. En outre, l'échange lent n'affecte pas les déplacements chimiques des deux états mais par contre ce processus réduit le temps de vie du spin dans un environnement donné. Comme dans le cas de l'échange rapide, les résonances s'élargissent sous l'effet de l'échange. Cette fois le terme R_{ex} pour pour l'état A par exemple est donnée par l'équation $R_{ex}^A = p_B k_{ex}$ (23) (Palmer et Massi 2006).

Dans le cas des échanges symétriques (avec des populations équivalentes pour les deux états associées à des vitesses k_1 et k_2 égales) ou d'un échange dans lequel la fraction minoritaire reste suffisamment grand comme dans l'exemple de la Figure 16 (k) (25%) les processus des mouvements lents peuvent être facilement identifiés sur les spectres RMN classiques à une ou à deux dimensions. Si l'état minoritaire est très faiblement peuplé (quelques %), une seule résonance, celle correspondant à l'état principal, est visible, mais elle élargie. Dans la littérature on parle d'état "invisible". De ce fait, les informations concernant la conformation minoritaire, comme la possibilité de déterminer les populations et les déplacements chimiques associés à cet état devient

impossible en utilisant des expériences de RMN classiques. Il est néanmoins possible aujourd'hui d'utiliser une catégorie particulière d'expériences (expériences de dispersion de relaxation) pour étudier les paramètres cinétiques, structuraux et thermodynamiques associés à ces échanges conformationnels de ce type. Dans ces expériences l'état minoritaire peu peuplé est étudié indirectement par l'influence qu'il exerce sur les vitesses de relaxation de l'état principal, les deux états étant connectés par l'échange conformationnel.

6.3 Stratégie d'étude des mouvements lents des ADN

Dans la littérature les noyaux les plus étudiés par les méthodes de dispersion de relaxation sont le ^{13}C et ^{15}N . Ces études ont d'abord été développées pour les protéines, leurs applications aux acides nucléiques ne s'étant faites que dans un deuxième temps. Le développement des méthodes de marquage isotopique des ADN durant les deux dernières décennies a rendu possible cette catégorie d'étude pour ce type de molécule. La plupart des séquences d'impulsions sont développées pour l'étude des groupements amides des liaisons peptidiques des protéines. Dans les acides nucléiques, le noyau idéal semble être l'azote du groupement imino, N1 des guanines et N3 des thymines. Néanmoins ce type de noyau a en fait été assez peu étudié dans le cas des acides nucléiques. En effet, les groupements iminos ne sont portés que par les thymines et les guanines ; aucune information n'est donc disponible ni sur les adénines et les cytosines, ni sur les sucres. De plus la perte des signaux iminos lorsque l'appariement est faible ou rompu dans des régions non canoniques rend ce type d'étude non utilisable notamment pour les ARN où les régions les plus intéressantes (boucles internes, boucles apicales) ne possèdent pas en général de protons iminos observables. Il faut noter néanmoins que ce dernier problème est moins critique pour les régions en double hélice (ADN ou ARN). Pour ces raisons, le noyau le plus utilisé pour l'étude des acides nucléiques est non pas l'azote mais le carbone qui permet des études complètes de la dynamique des bases aromatiques quelque soit le type de base (A, T/U; G ou C). On peut de plus étudier plusieurs carbones simultanément pour certaines bases (C6 et C5 pour les cytosines, C2 et C8 pour les adénines). Il est aussi possible de suivre les carbones des sucres ribose (ARN) et désoxyribose (ADN). La quantité d'information est donc clairement beaucoup plus importante que si les expériences n'étaient réalisées qu'avec l'azote. On peut évidemment, et c'est l'objet de certaines études, combiner les études sur les carbones et les azotes (Xue et al. 2015). Dans la suite de cette section nous allons

essentiellement nous concentrer sur les études de relaxation du ^{13}C et montrer les difficultés, les limites et les inconvénients des techniques d'études classiques.

6.3.1 Les expériences RMN de mesure des vitesses de relaxation

Les méthodes de mesure des vitesses R_1 et R_2 du carbone sont basées sur les méthodes conventionnelles hétéronucléaires de type HSQC ou TROSY (Pervushin, Wider, et Wüthrich 1998), ces dernières étant souvent préférées lors des études sur des macromolécules de taille élevée.

Les séquences de base des mesures des vitesses de relaxation R_1 , R_2 ont des formes similaires qui peuvent être divisées en 5 parties ou "bloc" (Figure 17) (A) transfert de polarisation, (B) relaxation, (C) évolution, (D) transfert inverse de polarisation et (E) acquisition.

Le premier "bloc" (A) de la série d'impulsions permet d'effectuer un transfert de type INEPT (*Insensitive Nucleus Enhanced Polarization Transfer*) qui consiste dans le transfert de la polarisation du proton au carbone et prépare l'aimantation carbone pour la mesure du R_1 (S_z) ou celle du R_2 ($S_{x/y}$) qui sera détaillée plus loin. Ensuite l'aimantation carbone évolue pendant la période dite de relaxation (B). En fait l'expérience est répétée plusieurs fois (10-20 expériences) avec des délais de relaxation durant lesquels on laisse l'aimantation longitudinale (R_1), transverse (R_2), ou celle opérant dans le repère tournant ($R_{1\rho}$), revenir à son état d'équilibre.

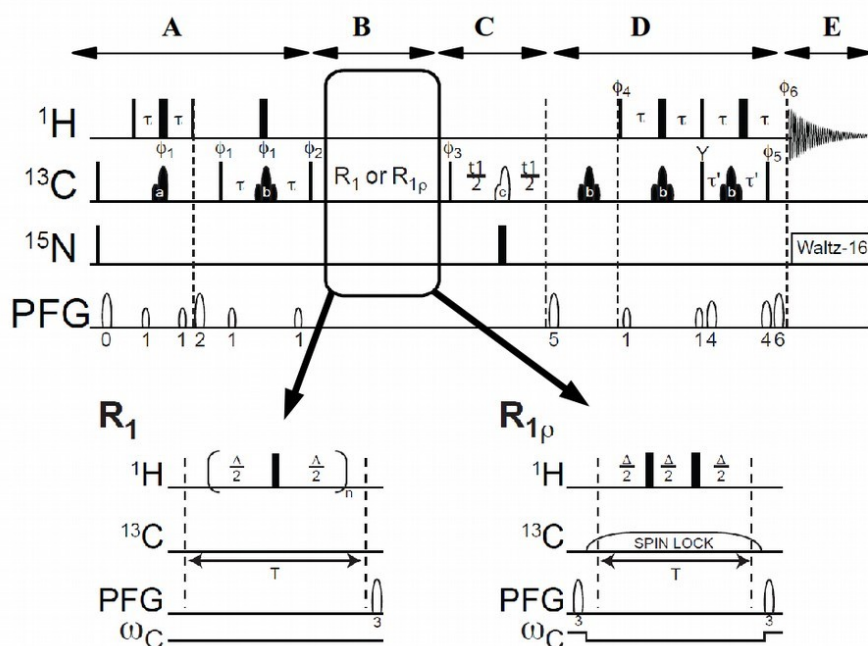


Figure 17 :Exemple de séquence d'impulsions 2D TROSY de mesure des relaxation ^{13}C R_1 ou $R_{1\rho}$ (Figure du haut). Les rectangles minces et épais indiquent les impulsions 90° et 180° respectivement, les impulsions non rectangulaires a, b et c représentent les "*shaped pulses*". Les deux figures du bas sont des représentations de la partie encadrée de la séquence du haut qui diffèrent en fonction de la fréquence de relaxation mesurée R_1 (à gauche) et $R_{1\rho}$ (à droite) (James G Kempf 2003) .

L'enregistrement de l'intensité de chaque pic croisé en fonction du délai de relaxation permet l'extraction de la vitesse de relaxation. L'introduction du découplage ^1H pendant cette période est essentielle et permet de supprimer les contributions indésirables. Le "*bloc*" suivant (C) l'aimantation évolue selon les déplacements chimiques permettant ainsi son marquage en fréquence et l'identification du résidu observé. Les couplages scalaires carbone-carbone (J_{CC}) présents pour les cytosines et les thymines (mais pas pour les guanines et les adénines) sont refocalisés en utilisant une évolution à temps constant. Le signal est acquis sur le canal proton (E) après le retour de l'aimantation sur ce dernier via le deuxième INEPT (retro-inept (D)).

Le "*bloc*" (B) de relaxation ^{13}C est évidemment différent selon que l'on souhaite mesurer R_1 , R_2 ou $R_{1\rho}$ dont je vais parler dans les paragraphes suivants. La méthode de type "inversion-retour" classique (Palmer, Kroenke, et Patrick Loria 2001) est utilisée pour la mesure de la vitesse de relaxation longitudinale (R_1). Au début du *bloc* de relaxation, l'aimantation est alignée selon l'axe-z et évolue durant des délais de relaxation variables. Le découplage ^1H est réalisé avec une série d'échos de spins séparés par de courts intervalles pour supprimer l'évolution sous les couplages ^1H - ^{13}C , les corrélations croisées entre les termes de type ^1H - ^{13}C dipolaire et anisotropie de déplacement chimique (CSA) ^{13}C ou les corrélations croisées entre les interactions dipolaires ^1H - ^{13}C et ^{13}C - ^{13}C . Lorsque ces dernières interactions croisées sont correctement supprimées, les intensités des pics résultant de ces expériences décroissent de façon monoexponentielle en fonction du délai de relaxation (Mulder et al. 1998). La vitesse de relaxation est donnée par la vitesse de décroissance de la courbe monoexponentielle résultante.

La relaxation transversale R_2 peut être mesurée en utilisant deux types d'expériences différentes: de type CPMG (*Carr-Purcell-Meiboom-Gill*)(Meiboom et Gill 1958) ou mesure de type $R_{1\rho}$. Dans le CPMG une série d'écho de spin est appliquée sur le canal du noyau d'intérêt $^{15}\text{N}/^{13}\text{C}$ durant la période de relaxation pour refocaliser l'aimantation transversale sur le plan x-y. Cette expérience est régulièrement utilisée pour la mesure de la relaxation des ^{15}N des groupements amides des protéines et les imino des acides nucléiques. Par contre l'utilisation de cette méthode n'est pas bien adaptée à l'étude de la relaxation des carbones des bases aromatiques et des sucres des acides

nucléiques notamment du fait des complications associées à la modulation de l'écho de spin, dues à la présence des couplages ^{13}C - ^{13}C (Yamazaki et al. 1994). Le choix du R1ρ pour la mesure de la relaxation des carbones des acides nucléiques apparaît la meilleure solution au vu des difficultés de réalisation de ce type d'expérience avec les séquences CPMG (Shajani et Varani 2008). Cette technique est également utilisée pour les études de relaxation du ^{13}C dans les protéines (Yamazaki et al., 1994).

La technique R1ρ est basée sur l'application de d'une impulsion RF à onde continue de type "*spin-lock*" de longue durée sur le carbone durant le bloc de relaxation (Figure 17 B), cette impulsion a pour effet de basculer l'aimantation dans le plan x-y selon un angle particulier (θ) (Figure 18). L'inclinaison effective de l'angle est directement relié à la force de champs RF appliqué (ω_1) et à l'écart entre la fréquence de l'impulsion et la résonance d'intérêt (Ω) avec la relation

$$\theta = \arctan\left(\frac{\omega_1}{\Omega}\right) \quad (24).$$

R1ρ dans le référentiel tournant contient les deux composantes longitudinale et transversale de la relaxation : $R_1\rho = R_1 \cos^2 \theta + R_2 \sin^2 \theta$ (25).

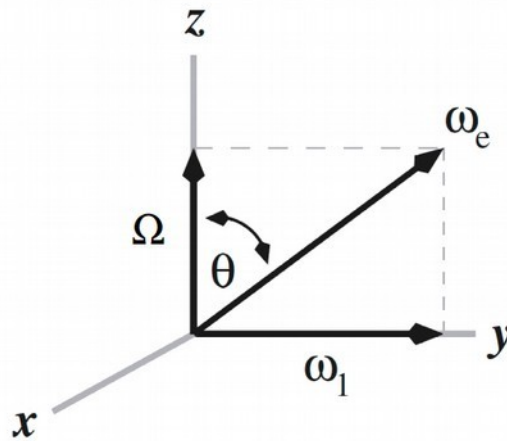


Figure 18 : Basculement de l'aimantation pendant la mesure R1ρ. Le champ de type "*spin lock*" est appliqué sur l'axe y avec une fréquence ω_1 , Ω est l'écart de la résonance d'intérêt par rapport à la porteuse "*Offset*". L'angle de basculement θ est donné par l'équation 24 et ω_e est la force de champ

effective ressentie localement par le spin.

Les expériences de type $R_1\rho$ produisent aussi des profils d'intensités de formes monoexponentielles à partir desquelles les vitesses de relaxation R_2 peuvent être calculées selon l'équation 25.

6.3.2 Expériences de dispersion de relaxation

Les expériences de dispersion-relaxation constituent un ensemble de méthodes qui permettent la quantification des phénomènes d'échange conformationnels lents. Elles reposent sur l'analyse des vitesses de relaxation transverses qui sont mesurées dans différentes conditions. La mesure de ces vitesses est réalisée avec le type de séquence décrit dans la section précédente (expériences CPMG ou $R_1\rho$).

Pour exposer le principe général de ces expériences de dispersion de relaxation, je vais prendre l'exemple schématique d'expériences utilisant les séquences de type CPMG de la Figure 19. Dans cette représentation la molécule (ou l'un des spins d'intérêt contenu dans cette molécule) est en échange conformationnel entre deux états A et B. Ces deux états conformationnels dont l'expérience cherche à sonder les propriétés, ont respectivement des fréquences de précession ω_A et ω_B différentes. En particulier, pendant le temps de relaxation les spins oscillent entre l'état A et l'état B, cet échange crée un déphasage de l'aimantation qui contribue positivement à la relaxation des spins selon l'équation (23). Comme il a été décrit auparavant, c'est la différence entre la vitesse d'échange (k_{ex}) et la différence en fréquence de précession entre les deux états ($\Delta\omega$) qui régissent la nature de l'échange et donc sa contribution à la relaxation transversale du spin.

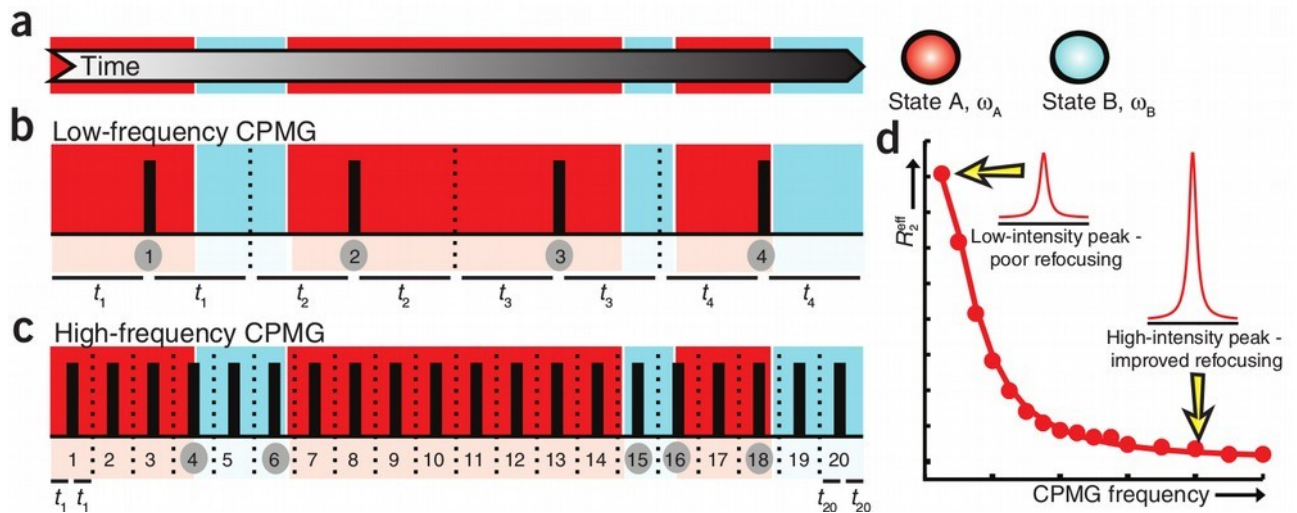


Figure 19 : Schéma simplifié d'une expérience CPMG de dispersion de relaxation. (a) La molécule est en échange stochastique entre deux états conformationnels A et B en fonction du temps qui sont dénotés par le rouge et le bleu respectivement ; (b et c) durant la trajectoire d'échange une série d'impulsions 180° sont appliquées pour refocaliser les aimantations ; (d) profil de dispersion de relaxation à partir duquel les paramètres d'échange peuvent être extraits (Baldwin et Kay 2009).

Il est bien entendu impossible d'agir sur la vitesse d'échange k_{ex} à moins de changer les conditions physique du milieu (température, etc ..). Par contre nous pouvons moduler les fréquences de résonance ($\Delta\omega$) en les refocalisant. La répétition des impulsions 180° de type écho de spin sur les spins pendant le temps de relaxation refocalise l'aimantation comme dans l'exemple de la Figure 20. L'effet de ces échos de spin peut être perçu en considérant par exemple le spin B dans la Figure 19; l'état B étant faiblement peuplé sa durée de vie est beaucoup moins importante de l'état A. L'application de séquences d'échos de spins avec des intervalles importants donne une faible probabilité (Figure 19b) de refocaliser les spins dans l'état B . Quand la fréquence des échos de spin devient importante, les probabilités de refocaliser les aimantations de l'état B deviennent importantes (partie (c) de la Figure 4). De même les fréquences élevés d'écho de spin refocalisent l'état A. L'augmentation de la fréquence du CPMG refocalise petit à petit les populations des états A et B jusqu'à ce que tout les spins soient alignés. En pratique, une série d'expériences CPMG est réalisée avec une augmentation de l'application des séquences d'écho de spin. Le résultat de ces expériences est donné dans la partie (d) de la Figure 19. Le graphique montre l'évolution de la vitesse de relaxation transversale en fonction de la fréquence de répétition des impulsions CPMG. L'accroissement de la fréquence des impulsions augmente la refocalisation de spins et donc diminue la contribution de l'échange conformationnel à la relaxation, et de ce fait la vitesse de relaxation

diminue.

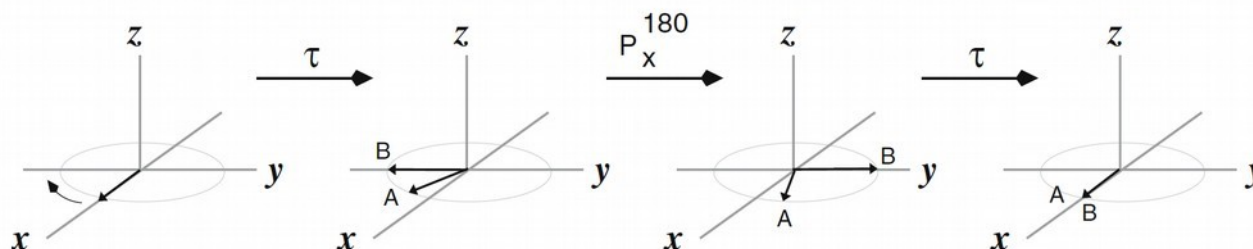


Figure 20 : Evolution de l'aimantation transversale pendant le temps du CPMG. La section la plus à gauche montre l'aimantation immédiatement après une impulsion 90° selon l'axe y . Après un temps τ les deux états A et B précessent à différentes fréquences, due à la différence des environnements locaux. Dans cet exemple la précession de l'état B est plus rapide que l'état A. L'impulsion 180° selon l'axe x va permuter les positions des spins A et B. Comme B a une vitesse supérieure à celle de A, après un temps τ les deux aimantations seront refocalisées (la dernière section à droite).

Dans la pratique, les expériences CPMG ne sont pas favorables pour des études de ^{13}C des acides nucléiques et des protéines du fait que les couplages ^{13}C - ^{13}C modulent l'enveloppe générée par l'écho de spin (Yamazaki et al. 1994). En fait, de telles études pourraient être réalisées sur des échantillons enrichis de façon sélective au niveau de carbones précis. De tels échantillons sont néanmoins encore très difficiles à obtenir. La plupart des études dans ce domaine (relaxation du ^{13}C des acides nucléiques) ont donc été effectuées avec des séquences de type R1 ρ . Cependant les expériences traditionnelles 2D R1 ρ ne sont sensibles qu'aux mouvements qui s'opèrent dans la gamme de temps de la microseconde, en raison des limites relatives aux forces de champ (ω_1) utilisés ($> 1000\text{Hz}$). Le résultat de cette limitation en ω_1 , est l'impossibilité de détecter des mouvements plus lents qu'à peu près $300\mu\text{s}$, la détection de mouvements plus lents de l'ordre de la milliseconde nécessite des valeurs de champ de ω_1 faibles. Cependant avec de faibles forces de champ ("spin-lock"), il devient difficile de supprimer l'évolution sous les couplages scalaires carbone-carbone. De plus, lorsque les champs sont faibles les effets "off-resonance" (les effets de l'angle de basculement des spins selon l'équation 24) deviennent critiques et il n'est plus possible d'aligner correctement l'aimantation des différents spins le long de leurs champs efficaces (Mulder et al. 1998) (Massi et al. 2004) (Korzhnev, Orekhov, et Kay 2005).

Des séquences d'impulsions utilisant des champs "spin lock" de faible valeur (c.à.d de l'ordre de

quelques centaines d'Hz), et prenant en compte ces différents problèmes ont été mises au point (Massi et al. 2004) (Korzhnev, Orekhov, et Kay 2005). Notamment une séquence d'impulsions R1ρ à une dimension a été mise au point par le groupe de Kay (Korzhnev, Orekhov, et Kay 2005). Dans ces conditions on s'éloigne des situations du type $\omega_e \gg k_{ex2}$ dans lequel k_{ex} est faible et il devient possible d'extraire les paramètres liés à l'échange tout en travaillant à un seul champ statique (alors que leur extraction avec des méthodes CPMG nécessite l'obtention de données à plusieurs champs). Cette méthode a permis l'étude des mouvements lents des groupements amides avec de faibles forces de champ allant jusqu'à 25Hz et l'extraction des paramètres tout en travaillant à un seul champ. Cette séquence d'impulsion, initialement mise au point pour les paires N-H amides dans les protéines a été adaptée à l'étude de la relaxation des carbones dans les acides nucléiques par le groupe d'Al-Hashimi en 2009 (Hansen et al. 2009) (Figure 21).

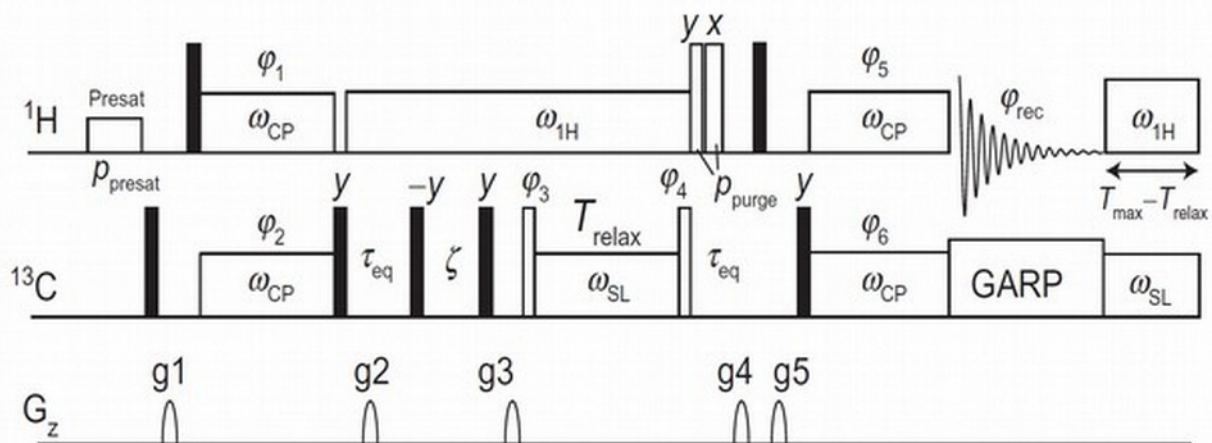


Figure 21 : Séquence d'impulsions "selective" ^{13}C R1ρ. Les rectangles noirs représentent les impulsions 90° dures, et les rectangles blanc les impulsions 90° sélectives. Les impulsions ω_{cp} , ω_{1H} et ω_{SL} sont détaillées dans le texte. Les gradients g1,g2,g3,g4 et g5 sont appliqués avec une durée d'une milliseconde (1ms) avec des profils SMSQ1.1000 et des amplitudes de 4.3, 8.9, 8.3, 9.7 et 6.1 G/cm respectivement. Le cycle de phase est le suivant : $\phi_1=\{8(y) 8(-y)\}$, $\phi_2=\{-x x\}$, $\phi_5=\{4(x) 4 (-x)\}$ et $\phi_6=\{2(x) 2(-x)\}$, $\phi_{req}=\{x -x -x x -x x x -x -x x x -x x -x -x x\}$. Dans les expériences "Off-resonance", $\Omega > 0$: $\phi_3=-y$, $\phi_4=y$ et $\Omega < 0$: $\phi_3=y$, $\phi_4=-y$ (Hansen et al. 2009)

Nous avons largement utilisé cette séquence dans ce travail. Cette méthode, de type 1D, repose sur un transfert initial sélectif d'aimantation par les méthodes de corrélation croisée (ω_{cp}) Hartmann-Hahn (Pelupessy et Chiarparin 2000). Chaque résonance (proton et carbone) est sélectivement excitée avec des forces de champ adéquates pour chacun des deux noyaux permettant un transfert Hartmann-Hahn hétéronucléaire sélectif. L'aimantation transférée sur le carbone, après un temps

(τ_{eq}) permettant aux spins en échange d'arriver à un état stationnaire, est alignée selon l'axe-z. Pendant le temps de relaxation (T_{relax}) une impulsion de type "*spin-lock*" est appliquée avec une force (ω_{SL}). L'aimantation carbone va donc s'aligner selon le champ efficace comme habituellement dans les expériences de type 2D R1 ρ . Le découplage du proton pendant le temps d'équilibre et le temps de relaxation est réalisé avec un découplage onde continue (ω_{1H}) de 10 ~ 12 KHz et permet d'éliminer les diverses contributions parasites (corrélations croisées entre les différents mécanismes de relaxation et *J-coupling* carbone-proton). Le délai ζ permet de filtrer les résonances indésirables les plus proches des carbones d'intérêt ($< 1.5 \omega_{cp}$). Le transfert inverse, toujours de type Hartmann-Hahn sélectif, est ensuite effectué, et un découplage d'environ 3.5KHz est appliqué sur le canal carbone pendant l'acquisition. Des impulsions à onde continue sont également appliquées sur les deux canaux 1H et ^{13}C à la fin de l'acquisition pour égaliser les temps de découplage 1H et le temps de l'application de "*spin-lock*" pendant les différents temps de relaxations. Ainsi, les temps d'irradiation sont identiques quelque soit le temps de relaxation utilisé.

La sélectivité permet la réalisation d'expériences "*On et Off-resonance*" qui sont effectuées séparément sur un seul noyau (Figure 22). L'intérêt de se focaliser sur une seule résonance lors de l'enregistrement de chaque spectre 1D est qu'il devient extrêmement simple d'aligner de façon appropriée l'aimantation selon le champ effectif. Les effets "*off-resonance*" ne posent donc plus de problème. En effet, l'utilisation d'expériences "*off-résonance*" est la stratégie utilisée dans ces expériences pour extraire finement les paramètres associés à l'échange conformationnel en utilisant un seul champ magnétique statique B_0 .

R1 ρ "*On-resonance*" à une dimension : Dans ces expériences, comme dans le CPMG, pendant le temps de relaxation, l'application de champs magnétiques radiofréquences (RF) de forces croissantes sur le spin d'intérêt, refocalise petit à petit les aimantations des deux états majoritaire et excité (Figure 22 C). Le profil simulé (Figure 22 E à gauche) montre la diminution de la contribution de l'échange (R_{ex}) à la relaxation du fait de l'alignement des aimantations des deux états.

R1 ρ "*Off-resonance*" à une dimension : cette expérience tend plutôt à augmenter la contribution de l'échange à la relaxation. Dans l'expérience "*Off-resonance*", pendant le temps de relaxation le "*spin-lock*" n'est pas appliqué à la fréquence de la résonance d'intérêt (l'état majoritaire), mais plutôt à d'autres fréquences variables (Ω). Quand l'impulsion est effectuée à la fréquence de l'état minoritaire, le déphasage des deux états majoritaire et minoritaire est maximal et la contribution de

l'échange à la relaxation sera très important. Le graphique de la Figure 22 E (à droite) montre des simulations des profils des expériences "Off-resonance". Il s'agit du suivi du $R_2 + R_{ex}$ en fonction des fréquences d'application des champs "spin-lock". Les profils montrent une courbe en cloche dont le sommet est centré sur la fréquence de l'état minoritaire. Ainsi, le déplacement chimique de l'état excité est résolu.

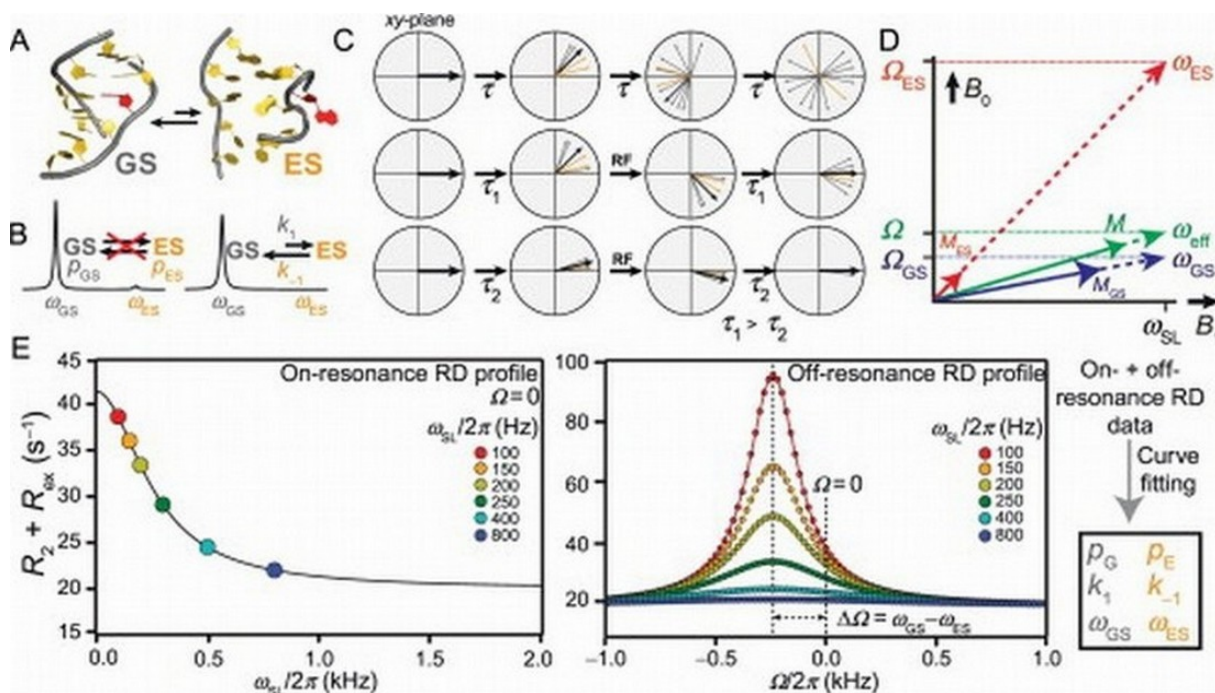


Figure 22 : Représentation schématique de la méthode de caractérisation des échanges conformationnels selon la méthode R1ρ (Xue et al. 2015). (a) exemple d'un équilibre asymétrique entre deux états conformationnels, état majoritaire (GS) et un état excité (ES). (b) spectre RMN 1D, en absence d'échange conformationnel (à gauche) et en présence d'échange conformationnel (à droite), dans ce dernier cas l'échange entraîne l'élargissement des raies de l'état majoritaire et minoritaire, ce dernier devient indétectable dans cet exemple. (c) évolution de l'aimantation transversale sur le plan x-y. Au début, l'aimantation est basculée sur le plan x-y; après un certain temps, soit τ , τ_1 ou τ_2 (du haut vers le bas), l'aimantation est déphasée dépendamment du temps (les vecteurs gris et jaunes représentent respectivement, l'état majoritaire et excité et le vecteur en noir la résultante du déphasage). Notons que $\tau_2 < \tau = \tau_1$. L'application du champ radiofréquence FR ("spin-lock"), dans l'exemple du milieu et celui du bas, aligne les aimantations selon leurs champs effectifs. ensuite, comme dans les expériences CPMG, l'aimantation est complètement refocalisée après un temps τ_2 ou partiellement en utilisant un temps plus long τ_1 . (d) le graphique représente les vecteurs des champs magnétiques effectifs ressentis par les noyaux dans les deux états conformationnels, majoritaire et excité, ω_{GS} et ω_{ES} respectivement en fonction de leur fréquences de résonances respectives Ω_{GS} et Ω_{ES} en ordonnée, et la force de champs appliquée (ω_{SL}) en abscisse. le vecteur M indique la somme des vecteurs des aimantations des deux états conformationnels, majoritaire (M_{GS}) et minoritaire (M_{ES}). Le ω_{eff} est le champ effectif à la fréquence moyenne des deux états (Ω). (e) simulations des profils de l'évolution de la relaxation transversale (R_2) auquel s'ajoute

la contribution de l'échange conformationnel (R_{ex}), en fonction de la force de champ appliquée dans les expériences *on-resonance* (à gauche), et en fonction de la fréquence à laquelle le RF est appliqué dans les expérience *off-resonance* (à droite).

Pour les échanges rapides ($k_{ex} > \Delta\omega$) l'expression de la vitesse de relaxation $R_1\rho$ est donnée par l'équation de Trott et Palmer (Trott et Palmer III 2002) :

$$R_1\rho = R_1 \cos^2 \theta + (R_2 + R_{ex}) \sin^2 \theta \quad (25) \quad \text{où} \quad R_{ex} = \frac{p_A p_B \Delta\omega^2 k_{ex}}{\omega_1^2 + k_{ex}^2} \quad (26) \quad \text{et } \theta \text{ est l'angle de}$$

basculement du champ effectif selon l'équation (24). Cette équation permet juste d'identifier l'existence de mouvements sur les carbones étudiés, mais elle n'est pas adéquate pour déterminer les paramètres de l'échange p_A , p_B et $\Delta\omega$ (Palmer et Massi 2006). Dans les cas où le système est asymétrique ($p_A \gg p_B$) qui est un cas fréquemment rencontré dans les études de dispersion de relaxation, les paramètres de mouvement peuvent être extraits de l'équation suivante, sans aucune hypothèse sur la vitesse de l'échange (Korzhnev, Orekhov, et Kay 2005) (Hansen et al. 2009) :

$$R_1\rho = R_1 \cos^2 \theta + R_2 \sin^2 \theta + \sin^2 \theta \frac{p_A p_B \Delta\omega^2 k_{ex}}{(\omega_A + \Delta\omega)^2 + \omega_1^2 + k_{ex}^2} \quad (27)$$

Cette méthode a montré son efficacité puisque des mouvements lents asymétriques de l'ordre de la milliseconde et des populations de l'état excité inférieures à 1 % ont pu être étudiés avec ces expériences. Cependant une attention particulière doit être prêtée à un autre phénomène qui peut contribuer à la relaxation. Il s'agit du phénomène de transfert Hartmann-Hahn durant le "*spin lock*" (HAHA "*matching*"). En effet ce phénomène intervient quand deux aimantations de deux spins de même nature et du même système, ici les carbones, sont alignées sur le même champ magnétique efficace pendant la durée du "*spin-lock*". Le fréquence efficace (ω_{eff}) d'un spin X et d'une fréquence Ω_X est donné $\omega_{eff} = \sqrt{\omega_1^2 + \Omega_X^2}$ (28) et ω_1 est directement relié à l'angle de l'impulsion ω_1 selon la relation (24). La contribution "Hartmann-Hahn *matching*" à la relaxation est donnée par la relation de Bax et Davis (Bax et Davis 1985).

$$A_{HAHA} = \left(1 + \left(\frac{\omega_{eff,I} - \omega_{eff,S}}{J_{IS} (1 + \cos(\theta_I - \theta_S)) / 2} \right)^2 \right)^{-1} \quad (29)$$

Selon l'angle de l'impulsion et l'écart par rapport à la porteuse (Ω) les champs efficaces de deux spins (I et S) peuvent se rapprocher voire devenir identiques et ainsi la contribution de ce phénomène à la relaxation devient maximale (100 %) puisqu'il est fonction de la différence entre les fréquences ω_{eff} de spins I et S. Cette contribution peut être évitée en choisissant soigneusement les

forces de champs et les angles d'impulsions pour lequel le A_{HAHA} est inférieur à 0.1 % après avoir déterminé ces contributions.

Un autre groupe, (Zhao, Hansen, et Zhang 2014) a publié une nouvelle méthode d'étude de la dispersion de relaxation des ^{13}C des ARNs à deux dimensions. Leurs travaux sont basés sur la nouvelle séquence d'impulsions *Chemical Exchange Saturation Transfer* (CEST) (Figure 23) mise au point par le groupe de Kay (Vallurupalli, Bouvignies, et Kay 2012) pour l'étude de la relaxation des groupements amides des protéines.

La séquence CEST se déroule de la façon suivante: au début de la période de relaxation, l'aimantation est alignée selon l'axe z et un champ RF (B_1) est appliquée sur le canal ^{15}N (Figure 23).

Quand l'impulsion B_1 appliquée est loin de la fréquence de résonance des pics majoritaire et minoritaire, l'intensité du pic majoritaire mesuré sur le spectre 2D n'est pas affectée par l'application de cette impulsion (l'intensité est égale à l'intensité du pic à $B_1 = 0\text{Hz}$). Cependant, quand B_1 est appliqué à la fréquence de résonance de l'état minoritaire, l'aimantation de cet état initialement sur l'axe z est basculée selon l'axe y. Ce basculement va induire une perte de cohérence entre les résonances en échange, ce qui diminue d'autant l'intensité du pic majoritaire. L'intérêt de la méthode est que l'aimantation sur laquelle l'impulsion est appliquée étant de type longitudinal, on peut utiliser des délais de relaxation beaucoup plus longs que quand les relaxations transversales sont utilisées, comme c'est le cas dans les expériences précédemment décrites du type CPMG et R1 ρ . Les vitesses de relaxation longitudinale et transverse peuvent différer d'un ordre de grandeur ou plus dans les protéines. La méthode est donc potentiellement beaucoup plus sensible que les méthodes de dispersion classique avec de meilleures possibilités pour quantifier les effets observés. Elle se révèle particulièrement utile pour l'étude de systèmes dans lesquels l'état minoritaire est très peu peuplé (Vallurupalli, Bouvignies, et Kay 2012).

Cette méthode a permis l'étude de mouvements très lents sur les protéines (jusqu'à 50 ms) et des systèmes dans lesquels la population de l'état minoritaire est d'environ 1 %. L'utilisation de cette méthode sur les ARN (Zhao, Hansen, et Zhang 2014) a permis l'étude des mouvements lents dans des régions confinées dans une gamme spectrale d'environ 4 ppm. Les auteurs de cette dernière publication, ont montré que cette méthode, permet de sonder les mouvements lents sur plusieurs noyaux en même temps (quelques ppm de gamme spectrale).

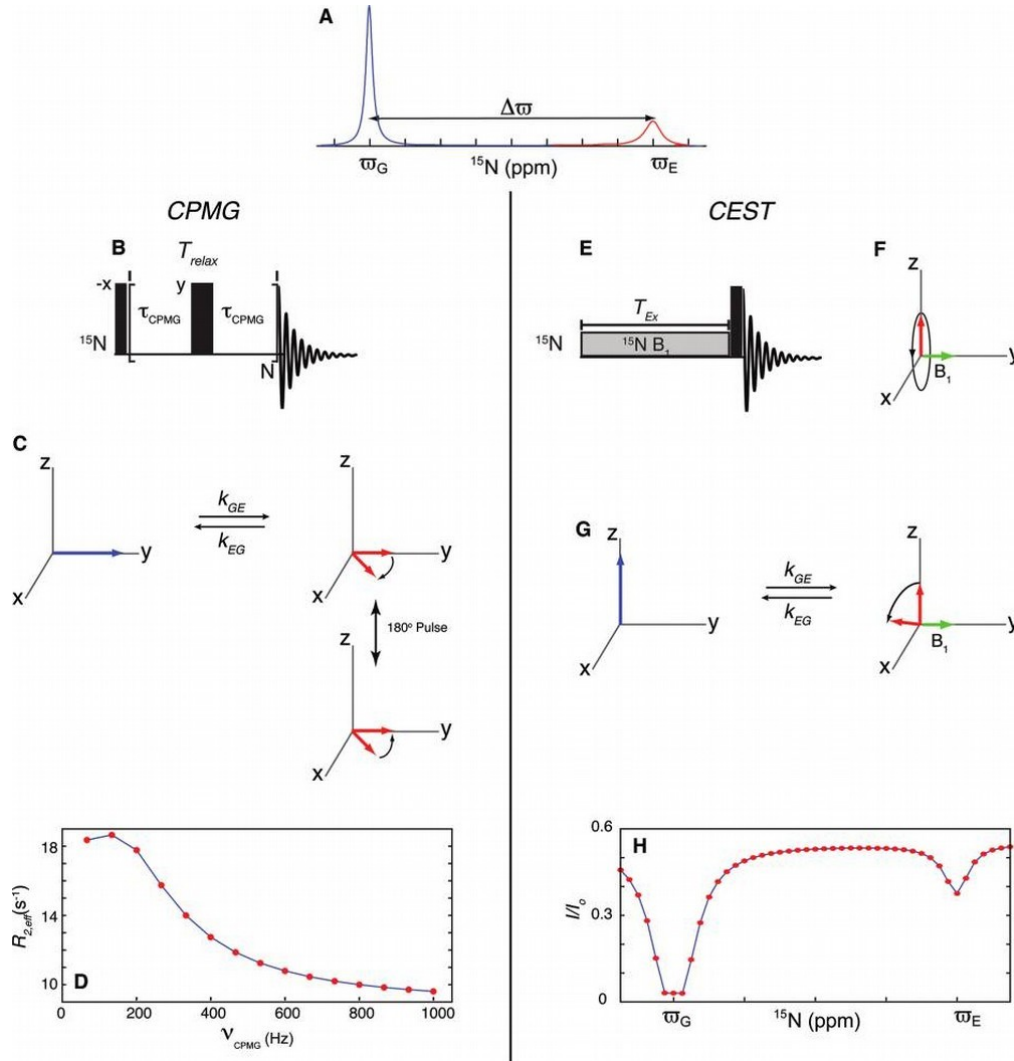


Figure 23: Comparaison entre le CPMG et le CEST (Vallurupalli, Bouvignies, et Kay 2012). (A) Spectre 1D ^{15}N d'un spin isolé s'échangeant entre deux états avec des déplacements chimiques ω_G (état majoritaire) et ω_E (état majoritaire). L'état minoritaire est montré dans le spectre juste pour illustration, mais en réalité il n'est pas observable dans le système d'intérêt asymétrique ($p_B \ll p_A$). (B) CPMG avec les impulsions 90° et 180° respectivement. Un nombre de répétition variable (N) d'écho de spin est appliqué pendant le temps de relaxation durant un délai constant T_{relax} . (C) illustration de l'aimantation de l'état majoritaire alignée selon l'axe y (en bleu), à droite l'état minoritaire en rouge. Le sens de la précession est ensuite inversée par l'impulsion 180° et l'aimantation refocalisée (D) Profil de relaxation transversale en fonction de la fréquence de répétition des impulsions 180. (E) Schéma illustrant le principe du CEST, une faible force de champ RF B_1 est appliquée sur le noyau azote. (F) quand le *spin-lock* (en vert) est à la fréquence de l'état minoritaire (en rouge) aligné selon l'axe z précesse autour de l'axe y . (G) à droite l'aimantation de l'état majoritaire (en bleu) est alignée selon l'axe z . L'application du *spin-lock* (en vert) à la fréquence de l'état minoritaire (en rouge) déphase ce dernier ce qui entraîne la baisse de l'intensité de l'état majoritaire. (H) Profil des intensités obtenu lors de la quantification de l'intensité de l'état majoritaire en fonction de position du champ faible d'irradiation. Le ratio I/I_0 , I est tracé après une période d'irradiation de durée T_{ex} et I_0 correspond à un $T_{\text{ex}}=0$. Les diminutions des intensités correspondent aux fréquences des états majoritaire (ω_G) et minoritaire (ω_E).

6.4 Les échanges conformationnels mis en évidence au sein des ADN-B

Les premiers travaux de dispersion de relaxation sur un ADN-B ont été réalisés en 2001 par l'équipe de James (Kojima et al. 2001). Dans ce travail, les chercheurs ont sondé l'existence de mouvements lents au niveau de l'adénine sur un dinucléotide CpA dans une séquence CAA.TTG elle même contenue dans un décacatère d'ADN enrichi à 13 % en ^{13}C et 98 % ^{15}N . L'utilisation des expériences R1 ρ 2D classiques ont permis à ces auteurs de rechercher l'existence de mouvements lents de l'ordre de la centaine de microseconde (limite du T1 ρ classique, *i.e* ; $\omega_1=1\text{kHZ}$). Cette étude a révélé l'existence de mouvements lents sur les profils de dispersion de relaxation du proton H8 et du carbone C8. Cependant la nature exacte de cet échange n'a pas été résolue. En 2011, les premiers travaux de dispersion de relaxation utilisant de faibles valeurs de "spin-lock" et donc permettant l'exploration de mouvements dans l'échelle de temps de la milliseconde avec les méthodes 1D R1 ρ ont été publiés par le groupe d'Al-Hashimi (Bothe, Lowenhaupt, et Al-Hashimi 2011). Il s'agissait de l'étude d'une séquence connue pour adopter la conformation Z. C'est une séquence riche en éléments GC qui était voisine d'une séquence TTTAT. L'étude de la dispersion de relaxation en abondance naturelle du plateau T:A précédent le motif GC répété 4 fois a révélé l'existence de mouvements lents dans le C8 de l'adénine et le C1' du désoxyribose de la thymine complémentaire. L'étude complémentaire par dichroïsme circulaire des mutants de cette séquence a montré l'importance de cette paire de bases T:A dans la transition B/Z. La même année, des transitions entre des appariements de type Watson-Crick (WC) et des appariements de type Hoogsteen (HG) dans des paires de bases localisés dans des ADN-B (Figure 24) ont été identifiés pour la première fois par la même équipe (Nikolova et al. 2011).

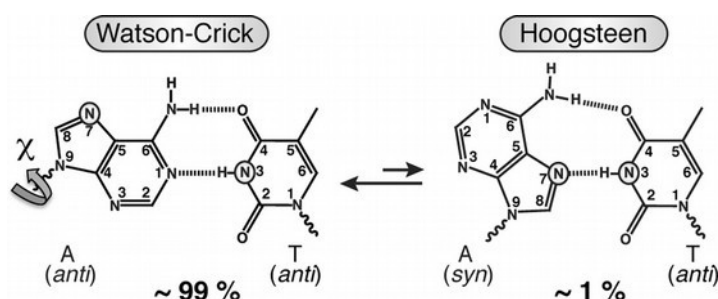


Figure 24 : Transition WC-HG (Nikolova et al. 2011). Une paire A-T Watson-Crick (à gauche) est majoritaire (99 %) mais peut s'échanger en une paire Hoogsteen (1 %). Le N1 accepteur dans l'adénine (WC) est remplacé par le N7 dans la paire (HG). L'angle de torsion χ entre les sucres et la base est en anti dans la paire WC et en syn dans la paire HG.

Des appariements Hoogsteen semblent pouvoir exister sous une forme minoritaire très peu peuplée (de l'ordre de 1%) sur certaines paires de bases se trouvant dans des contextes particuliers mais dans des séquences d'ADN tout à fait classiques. Ils peuvent être présents pendant des temps significativement longs (durée de vie supérieure à la milliseconde. De tels appariements, même transitoires, ont des propriétés électrostatiques et hydrophobes très différentes de celles de l'ADN classique. Étant donné qu'il y a un certain nombre d'exemples de facteurs de transcription (TBP, p53, etc...) qui reconnaissent spécifiquement des paires de type HG insérées dans des environnements classiques de type WC, la possibilité que ces états transitoires soient reconnus spécifiquement par des protéines constitue une hypothèse très attractive.

Les calculs de "*Density Functional Theory*" (DFT) ont montré que les déplacements chimiques des carbones de l'état excité obtenus par les expériences R1 ρ correspondent bien à des déplacements chimiques des bases en conformation Hoogsteen (Nikolova et al. 2011). Le même déplacement chimique spécifique de l'état excité (cad la forme Hoogsteen) a été retrouvé pour le carbone C8 lors de la méthylation en N1 de l'adénine (Nikolova, Gottardo, et Al-Hashimi 2012). Cette modification chimique force les bases dans la conformation Hoogsteen en déstabilisant l'appariement WC. De même, la substitution du N7 dans les bases adénines et guanines (la position N7 est accepteur de liaison hydrogènes dans les plateaux HG) (Voir Figure 24) en C7-H7, a pour effet de complètement éliminer les mouvements lents très probablement en empêchant la possibilité de réaliser des appariements de type Hoogsteen (Nikolova, Gottardo, et Al-Hashimi 2012). L'occurrence des transitions WC-HG a été mise en évidence (Alvey et al. 2014) dans différents contextes séquentiels avec des vitesses d'échanges variables d'une séquence à une autre ainsi que des fractions de l'état HG différentes. Cette dernière étude souligne le fait que ces transitions peuvent se produire dans une variété très importante de séquences.

En conclusion, bien que l'étude de la relaxation des protéines ait connu un grand progrès depuis une dizaine d'année, l'étude des mouvements lents des acides nucléiques s'est révélée difficile voire impossible en utilisant les méthodes classiques de dispersion de relaxation. De ce fait, cette branche de la RMN est demeurée très faiblement explorée et se limite à quelques publications. Le développement de nouvelles séquences d'impulsions telles que celles que nous avons présentées et utilisées rendent néanmoins maintenant possible l'étude de la dynamique de l'ADN à l'échelle de temps de la milliseconde. Elles permettent d'identifier des conformations transitoires et peu peuplées de l'ADN et de s'interroger sur le rôle potentiel que celles-ci peuvent jouer dans les processus biologiques.

7 Apport personnel

Dans le cadre de cette thèse, je me suis intéressé à la dynamique de l'ADN-B correspondant à des mouvements se produisant à deux échelles de temps, la pico-nanoseconde et la micro-milliseconde. Dans la continuité des sujets développés au laboratoire, j'ai d'abord étudié par RMN en solution les mouvements rapides BI \leftrightarrow BII des groupements phosphate sur quatre dodécamères non marqués. J'ai également réalisé des simulations de dynamique moléculaire sur ces mêmes dodécamères avec deux champs de force différents très récents, CHARMM36 et Parmbsc0_{OL1}. Dans un deuxième temps, j'ai sondé l'existence de mouvements plus lents par RMN, sur un des quatre dodécamères, marqué en $^{13}\text{C}/^{15}\text{N}$.

La séquence de ces 4 dodécamères n'a pas été choisie au hasard. Mis bout à bout, ces oligomères couvrent 39 paires de bases de la séquence 601, un DNA artificiel qui est connu pour sa capacité à former des nucléosomes (Thåström, Bingham, et Widom 2004). L'étude de nos dodécamères libres ouvre la possibilité de les comparer avec leurs équivalents au sein de la séquence 601 complexée avec le cœur d'histone. Une autre qualité de ces dodécamères est qu'ils contiennent une variété de dinucléotides suffisamment large pour envisager de tirer des conclusions solides sur l'effet de séquence sur la dynamique et sur la mécanique intrinsèque de l'ADN.

La première partie de ma thèse a été dédiée à l'étude de mouvements rapides, à partir d'expériences RMN dont les protocoles sont bien établis, et qui ont été réalisées avec Mlle Xiaoqian Xu, également en thèse au laboratoire. Nous avons d'abord extrait des spectres les déplacements chimiques de tous les phosphores et un grand nombre de distances. Dans un second temps, j'ai collecté un ensemble conséquent de constantes de couplages résiduels dipolaires (RDC). Ce très large jeu de données a ensuite été utilisé dans différents contextes.

Nous avons commencé par valider l'idée précédemment développée au laboratoire, à savoir qu'en solution, les mouvements des phosphates sont dépendants de la séquence dinucléotidique et prédictibles d'après l'échelle TRX, présentée en détail dans l'introduction. Nous avons ainsi pu utiliser en confiance cette échelle pour sonder les propriétés de différents ADN dont les affinités pour le cœur d'histones sont connues et formuler une hypothèse sur le mécanisme de reconnaissance de l'ADN par les histones. Ce travail a fait l'objet d'un article dans *Biochemistry* et sera exposé dans le chapitre 1.

Il nous a ensuite semblé intéressant de savoir si cet effet de séquence pouvait aussi être prédit par

des méthodes de simulations. C'est pourquoi j'ai testé les performances des champs de force CHARMM36 et Parmbcs0_{ε_{OL1}}, qui avaient été paramétrisés spécifiquement pour améliorer la représentation de la dynamique des phosphates. Les dynamiques moléculaires que j'ai réalisés en présence de solvant explicite, sur une durée totale de près de 5 μ s, ont montré que de grands progrès avaient été accomplis par rapport aux champs de force précédents, même s'il subsiste quelques défauts résiduels. Ce réalisme a permis de nous baser sur ces simulations pour accéder des informations dynamiques qui échappent à l'expérience. Ces résultats ont été décrits dans un article publié dans *Plos Computational Biology*, présenté dans le chapitre 2.

Enfin, nous avons abordé la mise en évidence de la mécanique intrinsèque de l'ADN en solution. Nous avons choisi d'étudier les relations entre trois types de données RMN, qui forment un ensemble cohérent caractérisant les dinucléotides : les déplacements chimiques du phosphore, les distances séquentielles (entre deux résidus successifs) et les Δ RDC, qui correspondent à la différence entre deux RDC successifs. De fait, les distances séquentielles d'une part, et les Δ RDC d'autre part, sont corrélés aux déplacements chimiques du phosphore. En prenant comme modèles des structures cristallographiques d'ADN-B à très haute résolution, j'ai ensuite déterminé à quels paramètres hélicoïdaux étaient sensibles les distances séquentielles et les Δ RDC. Mis en parallèle, ces résultats fournissent des évidences de la mécanique intrinsèque de l'ADN en solution, qui peut se résumer comme l'étroite inter-dépendance entre les conformations dynamiques des phosphates et des paramètres hélicoïdaux. Le chapitre 3 est consacré à cette thématique, sous forme d'un article qui soumis à *Nucleic Acids Research*.

Dans la dernière partie de ma thèse, j'aborde l'étude des mouvements lents de l'ADN. Nous avons remarqué des élargissements de raies sur une courte séquence riche en A:T dans un de nos dodécamères, qui laissaient soupçonner des mouvements lents. Sur cet oligomère enrichi en ^{13}C et ^{15}N , j'ai mis au point, calibré et réalisé des expériences de dispersion-relaxation de type 1D R1 ρ , décrites dans l'introduction. J'ai ainsi pu montrer l'existence d'un mouvement à l'échelle de la milliseconde qui n'avait jamais été décrit auparavant. Ces travaux sont présentés dans la chapitre 4. Ils sont préliminaires, mais ils constituent la base pour des études complémentaires qui seront menées au laboratoire.

CHAPITRE 1

L'échelle TRX et la formation du nucléosome

1 Introduction

Les résultats présentés ici reposent sur l'étude par RMN de quatre dodécamères, non marqués, reliés à la séquence 601, connue pour présenter des propriétés structurales et dynamiques -encore très mal connues- qui favorisent la formation du nucléosome. Les séquences nucléosomales couvrent entre 145 et 147 paires de bases, et sont donc trop longues pour pouvoir être étudiées en détail par RMN. Il faut donc les scinder. C'est ce que nous avons fait, en fragmentant 39 paires de base de la partie 5' de la séquence 601 en quatre dodécamères.

Un des principaux objectifs était de valider l'échelle TRX précédemment établie au laboratoire (Heddi et al. 2010). Comme nous l'avons présenté dans l'introduction générale, cette échelle associe des pourcentages de BII spécifiques (scores TRX) à chacun des dinucléotides de l'ADN-B. Les pourcentages de BII découlent de la traduction des δP selon une méthode empirique mise au point au laboratoire (Heddi et al. 2006) et que nous avons testée ici grâce aux mesures de 72 δP à trois températures (20, 30 et 40°C). En faisant l'hypothèse que la température ne devait pas influencer sur le δP correspondant à des états BI et BII également peuplés, nous avons trouvé que ce δP particulier était de 4.01 ppm. Traduit selon notre approche, cette valeur donne effectivement un pourcentage de BII très proche de 50 %. Une fois rassurés sur la qualité de notre méthode d'extrapolation, nous avons confronté les pourcentages de BII calculés à partir des δP collectés sur les quatre oligomères à ceux prédits par l'échelle TRX. La parfaite corrélation trouvée entre ces deux types de données nous a permis de valider l'échelle TRX, et aussi d'enrichir notre base de données avec les 72 δP mesurés dans ce travail.

Nous avons ensuite considéré un des effets structuraux associés aux états BI et BII. Les analyses les structures cristallographiques avaient montré que la densité de BII calculée sur des fenêtres

NpNpNpN.NpNpNpN (6 phosphates en tout) était couplée à la dimension des sillons, en particulier à la largeur du petit sillon qui peut varier de ~ 3.5 à 8 \AA (Oguey, Foloppe, et Hartmann 2010). Bien que cette variation puisse paraître assez large, il n'était pas évident de pouvoir la mettre en évidence en solution, sur des ADN non plus statiques mais dynamiques. En RMN, les NOE entre les H2 des adénines et les H1' de la base en 3' de la thymine complémentaire ((**H2**)ApN_i·(**H1'**)N_jpT, où N_i et N_j représentent n'importe quel nucléotide) renseignent sur la largeur du petit sillon (Chuprina et al. 1991). Dans nos dodécamères, il s'est avéré que ces NOE ne sont visibles que sur des régions riches en BI. Dans des zones riches en BII, ils ne sont pas vus alors que rien ne s'oppose à ce qu'ils puissent être observés (pas de superposition avec d'autres NOE, par exemple), ce qui signifie que les distances H2-H1' inter-brin sont supérieures à $5.5/6 \text{ \AA}$. Régions riches en BI et riches en BII correspondent donc bien à des petits sillons qui sont en moyenne, respectivement étroits et larges. En résumé, nous avons validé nos outils de quantification et de prédiction des pourcentages de BII en solution ainsi qu'enrichi et consolidé notre base de données. Nous avons également montré pour la première fois en solution le couplage qui existe entre la largeur du petit sillon et les conformations BI et BII.

La deuxième partie de l'article est consacrée à des séquences qui forment plus ou moins facilement des nucléosomes, en tenant compte des résultats obtenus dans la première partie. Dans un travail précédent (Heddi et al. 2010), il s'était avéré que les séquences de 145-147 paires de bases les plus aptes à former des nucléosomes étaient caractérisées par une périodicité TRX de 9 à 11 paires de bases, régions à haut score TRX alternant avec régions à faible score TRX, chacune de ces régions couvrant au moins 4 paires de bases (NpNpNpN.NpNpNpN, soit 3 dinucléotides). La taille de ces régions fait immédiatement penser à celle de la fenêtre modulant la taille du petit sillon (Oguey, Foloppe, et Hartmann 2010). Nous avons donc repris l'idée d'appliquer l'échelle TRX à un jeu de séquences nucléosomale, en considérant un score TRX dit « étendu », c'est à dire moyenné sur des fenêtres NpNpNpN.NpNpNpN. La corrélation entre les largeurs du petit sillon dans les structures cristallographiques du nucléosome et les scores TRX étendus est excellente pour les séquences de haute affinité. Cette corrélation s'affaiblit considérablement avec les séquences de faible affinité. Ces résultats suggèrent fortement que la taille des petits sillons de l'ADN libre, dans lesquels les arginines des histones vont s'ancrer pour stabiliser l'interaction histone-ADN, pourrait jouer un rôle majeur dans la formation du nucléosome. La lecture indirecte des séquences d'ADN par les histones serait ainsi liée à une prédisposition structurale et dynamique des sillons de l'ADN.

2 Article 1

NMR Studies of DNA Support the Role of Pre-Existing Minor Groove Variations in Nucleosome Indirect Readout

**Xiaoqian Xu, Akli Ben Imeddourene, Loussiné Zargarian, Nicolas Foloppe,
Olivier Mauffret, and Brigitte Hartmann**

Biochemistry 2014, 53, 5601–5612

NMR Studies of DNA Support the Role of Pre-Existing Minor Groove Variations in Nucleosome Indirect Readout

Xiaoqian Xu,^{†,‡} Akli Ben Imeddourene,^{†,§} Loussiné Zargarian,[†] Nicolas Foloppe,^{||} Olivier Mauffret,^{*,†} and Brigitte Hartmann^{*,†}

[†]LBPA, UMR 8113, ENS de Cachan CNRS, 61 avenue du Président Wilson, 94235 Cachan cedex, France

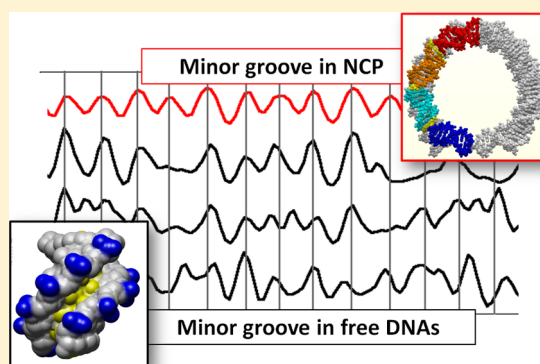
[‡]Department of Life Sciences, East China Normal University, 200062 Shanghai, People's Republic of China

[§]Pierre-and-Marie-Curie University, 75005 Paris, France

^{||}51 Natal Road, Cambridge CB1 3NY, United Kingdom

Supporting Information

ABSTRACT: We investigated how the intrinsic sequence-dependent properties probed via the phosphate linkages (BI \leftrightarrow BII equilibrium) influence the preferred shape of free DNA, and how this affects the nucleosome formation. First, this exploits NMR solution studies of four B-DNA dodecamers that together cover 39 base pairs of the 5' half of the sequence 601, of special interest for nucleosome formation. The results validate our previous prediction of a systematic, general sequence effect on the intrinsic backbone BII propensities. NMR provides new evidence that the backbone behavior is intimately coupled to the minor groove width. Second, application of the backbone behavior predictions to the full sequence 601 and other relevant sequences demonstrates that alternation of intrinsic low and high BII propensities, coupled to intrinsic narrow and wide minor grooves, largely coincides with the sinusoidal variations of the DNA minor groove width observed in crystallographic structures of the nucleosome. This correspondence is much poorer with low affinity sequences. Overall, the results indicate that nucleosome formation involves an indirect readout process implicating pre-existing DNA minor groove conformations. It also illustrates how the prediction of the intrinsic structural DNA behavior offers a powerful framework to gain explanatory insight on how proteins read DNA.



One of the most studied DNA–protein complexes is the nucleosome core particle (NCP), the fundamental building block of packaged DNA in eukaryotic cells. Nucleosomes offer a striking, and yet intriguing, case of indirect readout of DNA by proteins. On one hand, phosphodiester backbone atoms—and not base atoms—are involved in direct or water-mediated DNA–histone contacts.^{1–3} Therefore, the NCP formation cannot be guided by direct molecular recognition of the bases. On the other hand, the wrapping of DNA around the histone core requires marked structural DNA deformation, in particular, a bending of 30° per 10 base pairs accompanied by severe groove distortions.^{4–6} Thus, the way in which a DNA sequence can intrinsically and specifically modulate the shape and flexibility of the double-helix is thought to be an essential factor in NCP formation.^{6–9} Further support for the effect of DNA sequence lies in recurrent lexicographic signatures found in nucleosome positioning sequences.^{9–18} In these sequences, two categories of dinucleotides alternate out of phase with a period of ~10 base pairs (bp): ApA•TpT, TpA•TpA, ApT•ApT on one hand; and GpG•CpC, CpG•CpG, GpC•CpC, CpA•TpG on the other hand. The rationale of such alternations remains poorly understood. For instance, classifications of DNA flexibility

inferred from either X-ray DNA structures¹⁹ or stacking energies¹⁸ fall short of accounting for this periodicity. Both approaches categorize the dinucleotides in terms of flexible YpR, intermediate RpR•YpY, and stiff RpY steps whereas each alternating group identified in nucleosomal sequences is composed of a mixture of YpR, RpR•YpY, and RpY.

Several results suggest that, among the 145–147 bp of nucleosomal DNA, particular regions are more critical than others for forming the NCP. The SELEX method has been used to detect artificial DNA sequences with high affinity for the histone core in controlled conditions.²⁰ The highest-affinity sequence, called sequence 601,²⁰ is now widely used in biology research for its high ability to position nucleosomes, both *in vitro* and *in vivo*. The alignment of the highest-affinity SELEX sequences²⁰ has led to the identification of a conserved region covering 73 bp centered around the NCP dyad that illustrate the lexicographic periodicity mentioned above. The importance of this central part of positioning sequences has been further highlighted through analyses of nucleosome occupancy along

Received: April 26, 2014

Revised: August 4, 2014

Published: August 7, 2014

various genomes,²¹ and single-molecule experiments on the sequence 601.²² Interestingly, these findings echo earlier experiments showing that the nucleosome formation is initiated by the deposit of histones H3 and H4,^{23,24} which precisely contact the central part of nucleosomal DNA sequences. More specifically, the nucleosome assembly is highly sensitive to the sequence of the two segments located symmetrically at 1.5 double-helix turns from the dyad (SHL \pm 1.5, SHL stands for SuperHelix Location).²⁵ In NCPs crystallized with the human α -satellite sequence,^{1,3,8} or sequence 601^{26,27} and its derivatives,^{5,28,29} SHL \pm 1.5 segments are characterized by a spectacular, extremely narrow minor groove; they are implicated in a unique network of interactions involving van der Waals contacts with hydrophobic side chains, Leu65 in H3 and Pro32 in H4, and electrostatic interactions with Arg63 in H3. Sequence 601 contains SHL \pm 1.5 elements composed of T•A base pairs that, when they are associated with a marked narrow minor groove, generate a strong electronegative potential which strengthens the interaction with positively charged amino acids.³⁰ In NCPs, the conjunction of T•A-rich segments and narrow minor groove thus reinforces the interaction with Arg63 at SHL \pm 1.5.³¹ These observations suggest that T•A-rich sequences associated with pre-existing narrow minor groove could promote the nucleosome formation, strengthening the idea of a role of the sequence-dependent intrinsic DNA properties in indirect readout.

Ideally, identifying structural signatures of DNA sequences that mediate indirect readout during nucleosome formation requires comparison of free, unbound DNA to the same DNA in formed nucleosome. However, practically, deciphering free DNA structural properties in solution remains a challenge.^{32,33} That is because free DNA usually deviates only in subtle ways from a regular B-double helix, and this small deviations are difficult to ascertain experimentally or computationally.

Recently, the modulation of the structure of free B-DNA has been characterized in solution by considering the phosphate group BI↔BII equilibrium, which is sequence-dependent and closely related to the overall shape of B-DNA.^{33,34} In B-DNA in solution, the phosphate groups oscillate between two states, defined by the torsion angles ϵ and ζ , *trans/g-* in BI (ϵ - ζ \sim -90°) and *g-/trans* in BII (ϵ - ζ \sim $+90^\circ$) (Figure 1).^{35,36} This conformational transition occurs on a nanosecond time scale.³⁷ The BII propensity of each phosphate can be determined in

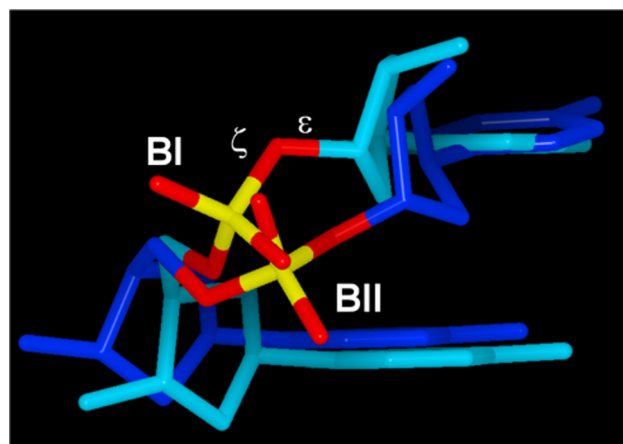


Figure 1. BI and BII conformations. Illustration of the BI (bases and sugars in light blue, (ϵ - ζ) = -90°) and BII (bases and sugars in dark blue, (ϵ - ζ) = $+90^\circ$) phosphate linkage conformations with a CpC dinucleotide.

solution by NMR, from the ^{31}P chemical shifts (δP).^{38,39} Importantly, each of the 16 B-DNA dinucleotides is characterized by a specific δP value.³⁴ Hence, specific BII propensity scores were proposed to describe and predict the backbone behavior at the dinucleotide level.³⁴

In addition, analyses of X-ray structures and molecular modeling studies showed that the conformations of the phosphates of a complementary dinucleotide, BI•BI, BI•BII (and its counterpart BII•BI) or BII•BII, are associated with distinctive values of the local, inter-base-pair parameters of twist, roll, and slide.^{38,40–45} A recent modeling study mentioned that rise, tilt, and shift are also affected by the backbone states.⁴⁶ Twist, roll, and slide are the major factors associated with B-DNA structure variability.¹⁹ Dinucleotides in BI are associated with low twist, null, or positive roll, and negative slide (BI profile). Dinucleotides in BII are characterized by high twist, negative roll, and positive slide (BII profile). Some dinucleotides are essentially confined to the BI state and the associated local helicoidal BI profile. In contrast, dinucleotides with a significant BII population explore a larger conformational landscape corresponding to both BI and BII profiles.

In B-DNA X-ray structures, the groove shape within tetramer segments is also related to the BII density, via a cumulative effect on base-pair displacement (X-disp).^{47–49} BI-rich tetramers present a typical narrow minor groove, while an accumulation of BII phosphate groups favors a more open and accessible minor groove (Figure 2).^{47,48} This intrinsic DNA

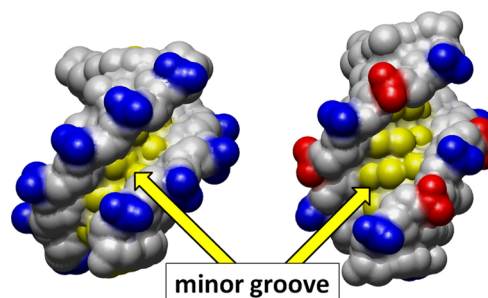


Figure 2. Illustration of the structural relationship between BI- or BII-rich regions and the DNA minor groove width. The GpG•CpC dinucleotides contain two facing phosphate groups either in BI (left, BI phosphate groups in blue) or in BII (right, BII phosphate groups in red). The yellow arrows emphasize the minor grooves, either narrow (left) or wider (right). The base atoms constituting the minor groove floor are in yellow. The structures were extracted from DNA crystal structures with PDB entries 1EHV and 3GGI.

property is exploited by minor groove binding proteins.^{47,50} Thus, the BI and BII states correlate with the sequence-dependent intrinsic DNA variability in structure, which in turn influences the molecular recognition of DNA via indirect readout.

In sum, δP s and BII propensities are useful and convenient surrogate reporters to inform about the DNA shape and its sequence dependent variation in solution. In reference to the relationship observed in X-ray structures between BI/BII states, Twist, Roll, and X-disp, the so-called TRX scale was proposed to characterize the 10 complementary dinucleotides by their BII propensities, inferred from a large data set of δP s.³⁴

On the basis of the above findings, we investigate here how the intrinsic DNA properties influence the sequence-dependent formation of the NCP. We used NMR to study four constituent

overlapping dodecamers of the sequence 601. Together, these oligomers cover a nonoverlapping 39 bp segment in the 5' half of sequence 601 (from −4 to −0.1 in terms of SHL) (Figure 3). Oligomer 3 is centered on the TTAAA element situated at $\text{SHL} \pm 1.5$ in the NCP, which is potentially a stringent singular positioning signal, as explained above.

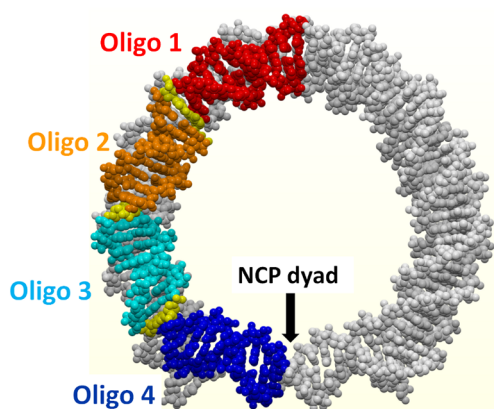


Figure 3. Location of the four studied dodecamers in Nucleosome Core Particle (NCP). Four dodecamers overlapping by three bases (Oligo 1–4, sequences given in Table 1) were studied in their free (unbound) state by NMR. The juxtaposition of the four oligomers recomposes a sequence that corresponds to part of the 5' half of the DNA sequence 601 (from −4 to −0.1 in terms of SHL). Here, the four dodecamers are positioned on the structure of the sequence 601 within NCP (PDB entry 3MVD) in which the DNA is wrapped twice around the histone core (not shown in this figure). Oligo 1 is in red, Oligo 2 in orange, Oligo 3 in light blue, and Oligo 4 in dark blue. The base pairs in yellow correspond to the overlapping bases in the free studied dodecamers.

We start the present study with methodological elements. First, examination of the temperature-dependence of ^{31}P chemical shifts enabled us to further validate the method relating BII percentages to ^{31}P chemical shifts.³⁸ Second, the new BII percentages inferred from the ^{31}P chemical shifts measured for the four oligomers provided a stringent test for the previously proposed prediction of BII scores from DNA sequence.³⁴ The predicted BII propensities were found in excellent agreement with the new BII percentages gathered on the four oligomers, proving that the BII propensities can be predicted from sequence alone. Third, the NMR data also confirmed that BII density is coupled to minor groove width as previously inferred from X-ray structure analysis.⁴⁷

We then investigated the relationship between minor groove width variations in the NCP and their counterparts in free

DNA, which are represented by backbone intrinsic BII propensity. One of the most interesting features of DNA in the nucleosome is the periodic (“sinusoidal”) alternation of wide and narrow minor grooves, located outward and inward from the histone core, respectively.^{4–6} The location of minor groove width minima and maxima is remarkably maintained over 35 X-ray nucleosome structures,⁶ while the distribution of slide and roll are quite variable along the DNA sequences.⁴ By extending the prediction of intrinsic BII propensities to the whole sequence 601 and five other artificial sequences of high or low affinity for the histone core, the present work uncovers a parallel between the intrinsic properties of DNA free in solution and the distortions of DNA observed in X-ray structures of the NCPs. This provides new insights into the ability of sequences at forming NCPs, and suggests an indirect readout of the DNA groove shape.

MATERIALS AND METHODS

DNA Sequences. Four oligodeoxyribonucleotides of 12 base pairs (bp) were studied by NMR. Their sequences are given in Table 1. The sequences of these dodecamers overlap by three bases. These overlaps make it possible to splice the four oligomers without considering the terminal base pairs, subject to end effects. For instance, Oligo 1 ends with $\text{G}_{10}\text{C}_{11}\text{T}_{12}$ and Oligo 2 begins with the same motif, $\text{G}_1\text{C}_2\text{T}_3$. The NMR data of the terminal dinucleotides $\text{C}_{11}\text{T}_{12}$ and G_1C_2 were discarded for the analyses while those collected for the penultimate $\text{G}_{10}\text{C}_{11}$ and C_2T_3 steps were considered.

The four juxtaposed dodecamers recompose a 39 bp sequence (Table 1) that corresponds to part of the non-palindromic sequence 601 of 146 bp (Table 1). The sequence 601 was selected for its very high-affinity for association with the histone octamer, as detected in SELEX experiments.²⁰ The center of the sequence 601 defines two 73 bp halves. In NCP structures, this DNA center corresponds to the pseudo-twofold axis of symmetry, the NCP dyad. According to the conventions suggested for the description of the first X-ray structure of NCP,³ the rotational orientation of the DNA is defined relative to the DNA center (SuperHelix Location zero, or SHL0). For each successive turn, the location number increases in the 3' half up to $\text{SHL}+7$, and decreases in the 5' half down to $\text{SHL}-7$. The four dodecamers studied by NMR corresponds to the segment from $\text{SHL}-4$ to $\text{SHL}-0.1$ in NCP (Figure 3).

Sample Preparation. Oligomers were synthesized by Eurogentec Inc. (Belgium). The sample was dissolved in an aqueous sodium phosphate buffer corresponding to an ionic strength of 0.1 (mol/L) with 0.1 mM EDTA, at pH 6.5. The duplexes were prepared by mixing the two complementary strands in a 1:1 ratio in 450 μL H_2O and 50 μL D_2O for studies

Table 1. Sequences of the Four Contiguous DNA Dodecamers Constituents of the Sequence 601^a

Oligo 1	5'-TCGTAGCAAGCT-3'•5'-AGCTTGCTACGA-3'
Oligo 2	5'-GCTCTAGCACCG-3'•5'-CGGTGCTAGAGC-3'
Oligo 3	5'-CCGCTTAAACGC-3'•5'-GCGTTTAAGCGG-3'
Oligo 4	5'-CGCACGTACGCG-3'•5'-CGCGTACGTGCG-3'
39 bp	$\text{T}_1\text{C}_2\text{G}_3\text{T}_4\text{A}_5\text{G}_6\text{C}_7\text{A}_8\text{A}_9\text{G}_{10}\text{C}_{11}\text{T}_{12}\text{C}_{13}\text{T}_{14}\text{A}_{15}\text{G}_{16}\text{C}_{17}\text{A}_{18}\text{C}_{19}\text{G}_{20}\text{C}_{21}\text{C}_{22}\text{T}_{23}\text{T}_{24}\text{A}_{25}\text{A}_{26}\text{A}_{27}\text{C}_{28}\text{G}_{29}\text{C}_{30}\text{A}_{31}\text{C}_{32}\text{G}_{33}\text{T}_{34}\text{A}_{35}\text{C}_{36}\text{G}_{37}\text{C}_{38}\text{G}_{39}$
Sequence 601	TGGAGAAATCCCGGTGCTAAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGC TGTCCTCCCGCGTTT-TAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGT

^aFour oligomers were studied free in solution by NMR. Their sequences overlap end to end by three bases (underlined). These overlaps make it possible to juxtapose the central parts of the four oligomers without considering the terminal base pairs (subject to end effects) and to recompose a continuous 39 bp sequence corresponding to part of the sequence 601. The 5' → 3' sequence of this 39 bp segment and its numbering are shown below the sequences of the oligomers. The last row corresponds to the 5' → 3' sequence 601 of 146 bp; its center, indicated by a double bar, defines two 73 bp halves of a same strand; the 39 bp fragment is underlined (see also Figure 3).

Table 2. Predicted BII Percentages of Complementary Dinucleotides in B-DNA^a

		sequence	BII _{i,av} %•BII _{j,av} %	BII _{i,j,av} %	SD BII _{i,j,av} %
complementary dinucleotides	YpR•YpR	CpG•CpG	43•43	43	16
		CpA•TpG	52•31	42	18
		TpA•TpA	14•14	14	7
	RpR•YpY	GpG•CpC	47•37	42	11
		GpA•TpC	33•11	22	15
		ApG•CpT	18•0	9	15
	RpY•RpY	ApA•TpT	11•0	5	12
		GpC•GpC	25•25	25	11
		ApC•GpT	8•0	4	13
		ApT•ApT	0•0	0	14
3'- and 5'-neighbor effect	CpG•CpG	YCpGR•YCpGR	59•59	59	19
		YCpGY•RCpGR	51•51	51	14
		RCpGY•RCpGY	38•38	38	17
	CpA•TpG	YCpAR•YTpGR	77•42	59	8
		YCpAY•RTpGR	59•21	40	11
		RCpAR•YTpGY	33•28	30	18
		RCpAY•RTpGY	36•19	27	17

^aThe 10 complementary dinucleotides Np_iN•Np_jN are characterized by BII_{i,j,av}%, which are the half-sums of the individual BII_{i,av}% and BII_{j,av}% of two facing phosphates. These BII percentages were previously extracted from a data set of 323 δ P.³⁴ The BII percentages of CpG•CpG and CpA•TpG are also given considering the effect of their 3'- and 5'-nearest neighbors, expressed in terms of purine (R) and pyrimidine (Y). The standard deviations (SD) given for BII_{i,j,av}% were inferred from the δ P variabilities (see Materials and Methods).

of exchangeable protons. For studies of nonexchangeable protons, the duplexes were lyophilized three times and dissolved in 500 μ L of 99.9% D₂O. The final concentration of double strand oligomers is 1–1.5 mM.

NMR Spectroscopy. All NMR spectra were recorded on a Bruker Avance spectrometer operating at a proton frequency of 500 MHz and at a phosphorus frequency of 202 MHz with a 5 mm gradient indirect probe. All spectra were processed with NMRpipe⁵¹ and analyzed with Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

One-dimensional ¹H spectra collected from 5 to 60 °C with a 5 °C step enabled us to check that the duplexes were stable over this range of temperatures. All the imino protons were clearly observable at 50 °C, except those of the terminal base pairs.

¹H NMR studies were performed at 10, 20, and 30 °C. 2D NOESY spectra were recorded using mixing times of 100, 200, and 300 ms for exchangeable protons and 80, 100, 200, 300, and 400 ms for nonexchangeable protons. MLEV-17 TOCSY experiments were run using mixing time of 120 ms. The 2D NOESY and MLEV-17 TOCSY were recorded under the following experimental conditions: 2048 data points and 512 t1 increments with spectral widths of 10 000 Hz for exchangeable protons and 5000 Hz for nonexchangeable protons. The water signal was suppressed with a WATERGATE sequence.⁵² NOE distances were extracted from cross-peaks with particular care after visual inspection of the build-up rates and using the distance extrapolation method to correct the spin-diffusion effects.⁵³ Distances were normalized using the cytosine H5–H6 proton pairs ($r = 2.5$ Å) apart from those involving the CH₃ group, for which the reference was the average H6–CH₃ distance ($r = 2.9$ Å). The experimental error on distances is estimated to 10% of the measured distances.

¹H–³¹P HETCOR⁵⁴ experiments were run at 20, 30, and 40 °C. The spectra width was 2500 Hz in the ¹H dimension and 810 Hz in the ³¹P dimension. Data were recorded with 2048 points in the ¹H dimension and 256 increments in the ³¹P dimension. ³¹P chemical shifts were referenced relative to

internal trimethyl phosphate. Uncertainty of ³¹P chemical shift measured in solution is estimated to ± 0.02 ppm.

Apart from the section ³¹P Chemical Shifts to Characterize the DNA Backbone Behavior, the NMR data presented here were collected at 20 °C.

All the NMR data are available in the Biological Magnetic Resonance Bank (BMRB) entry 19222.

BII Propensities. BII percentages (BII%) of the phosphate linkages along the four dodecamers were inferred from the δ P_s measured here at 20 °C using the equation $BII(\%) = 143 \delta P + 621$ (δ P referenced to trimethyl phosphate).³⁸ BII% were used in the section ³¹P Chemical Shifts to Characterize the DNA Backbone Behavior.

In the section Validation of the DNA Sequence Effect on the Backbone Behavior with the Dodecamers, we considered the BII propensity of the complementary dinucleotides Np_iN•Np_jN defined as the half-sum of the individual BII percentages of the facing phosphates p_i and p_j: $BII_{i,j}(\%) = (BII_i(\%) + BII_j(\%))/2$. In Table 2, the 10 complementary dinucleotides Np_iN•Np_jN are characterized by specific BII_{i,j}% previously published.³⁴ The corresponding standard deviations were inferred from the δ P variabilities between various instances of a given step, which are on average ± 0.10 ppm. These variances dominate the δ P experimental errors (± 0.02 ppm) that were therefore neglected.

In Relation between BII Propensities and Minor Groove Width Validated in Solution with NMR and Implications for the Preferential Interaction of DNA Sequences with the Histone Core, we defined an extended BII score, BII_{i,j,ext}%, in order to take into account the effect of the 5' and 3' neighbors on the base pair displacements of Np_iN•Np_jN, which are related to groove dimensions.⁴⁷ Thus, the complementary dinucleotides Np_iN•Np_jN are characterized by BII_{i,j,ext}%, the averaged sum of the individual BII_{i,j}% of the three complementary dinucleotides in the tetramer Np_{i-1}Np_iNp_{i+1}N•Np_{j-1}Np_jNp_{j+1}N: $BII_{i,j,ext}(\%) = (BII_{i-1,j+1}(\%) + BII_{i,j}(\%) + BII_{i+1,j-1}(\%))/3$.

Crystallographic NCP Structures. In the context of this study, we examined the three NCP structures crystallized with sequence 601 of either 145 bp (PDB entries 3ZL0 and 3ZL1,

resolution of 2.5 Å²⁷) or 146 bp (PDB entries 3MVD, resolution of 2.9 Å²⁶). The additional base in 3MVD is located at the end of the sequence (...CGAT-3' instead of ...GAT-3' in 3LZ0 and 3LZ1). The 3MVD structure contains the chromosome condensation regulator RCC1 that contacts three phosphate groups without disturbing the overall DNA structure, according to our analyses. Analyses of NCP structures were carried out using Curves+.⁵⁵

Graphics. Graphical representations were prepared with Yasara (<http://www.yasara.org/>).

RESULTS AND DISCUSSION

³¹P Chemical Shifts to Characterize the DNA Backbone Behavior. The ³¹P chemical shift (δP) values reflect the two-state BI \leftrightarrow BII equilibrium (Figure 1). High- and downfield shifted δP s correspond to BI- and BII-rich phosphate groups, respectively.^{36,37} The four oligomers studied here (Table 1) correspond to 72 dinucleotide steps, excluding the terminal steps. All 72 δP s were assigned and measured at 20, 30, and 40 °C (Table S1 in Supporting Information). They cover a range typical of B-DNA, from -3.80 to -4.51 ppm. Examples of ¹H-³¹P HETCOR spectra are shown in Figure 4.

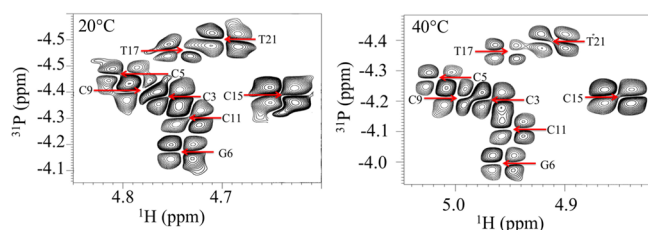


Figure 4. Representative ¹H-³¹P HETCOR spectra for δP determination. The ¹H-³¹P HETCOR spectra show a section of the H3'-P region obtained for Oligo 4 at 20 °C (left panel) and 40 °C (right panel). The red arrows indicate the cross-peak centers.

Two empirical relations were proposed to translate δP s in terms of BII percentages,^{38,39} using eq 1³⁸ or 2 at 20 °C.³⁹

$$\text{BII}\% = 143\delta P + 621 \quad (1)$$

$$\text{BII}\% = 114.84\delta P + 515.62 \quad (2)$$

These equations were obtained with two different methods, which both assume implicitly³⁸ or explicitly³⁹ that δP of pure BI and pure BII states are sequence-independent. This assumption is very difficult to test by NMR, since, to our knowledge, no other experimental method can identify which dinucleotide would be strictly confined in only one state. A recent interesting computational study, combining molecular dynamics simulations on the Drew-Dickerson dodecamer and quantum mechanical calculations of δP , suggested that the backbone angle values are sensitive to the dinucleotide sequence; according to DFT calculations on DNA fragments, these sequence dependent variations would affect the δP values of pure BI and BII states.⁵⁶ Although eqs 1 and 2 do not yet integrate such refinement, the overall consistency of the results presented below indicates that they are robust for estimating BII%.

Examining the behavior of the phosphate linkages as a function of temperature enabled to estimate the δP value that does not vary with temperature. This particular value is expected to be the signature of equal populations of BI and BII (BII% = 50).^{57,58} The temperature study investigated the influence of a moderate increase of temperature on the

BI \leftrightarrow BII equilibrium by collecting δP s at 20, 30, and 40 °C, on integrally preserved double helices. The δP values changed by 0.043 ± 0.022 ppm on average between 20 and 40 °C (typical changes in Figure S1 in Supporting Information). Such slight but significant variations of the same magnitude were previously observed on other oligomers.^{39,58}

The variation of δP between 20 and 40 °C ($\Delta\delta P$) is negatively correlated to the reference δP values at 20 °C (Figure 5).

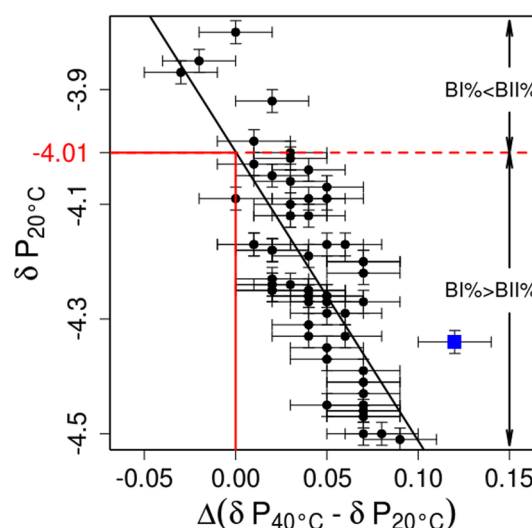


Figure 5. Relationship between the amplitude of ³¹P chemical shift changes according to temperature and the down- or high-field character of phosphate linkages. The ³¹P chemical shift change between 20 and 40 °C ($\Delta(\delta P_{40^\circ} - \delta P_{20^\circ})$, ppm) is linearly correlated to δP measured at 20 °C (δP_{20° , ppm). The phosphate linkage of ApA (blue square) in the TTAA segment of Oligo 3 has an unusual behavior (see the text). The black line corresponds to the linear fit of the data, discarding δP of ApA in TTAA. $\Delta(\delta P_{40^\circ} - \delta P_{20^\circ}) = 0$ (marked in red) theoretically corresponds to a BI/BII ratio of 1. Accordingly, $\delta P_{\text{BI/BII}=1} = -4.01$ ppm. $\delta P > \delta P_{\text{BI/BII}=1}$ and $\delta P < \delta P_{\text{BI/BII}=1}$ correspond to a majority of BII and BI conformers, respectively. The bars are the experimental errors.

Positive and negative $\Delta\delta P$ s correspond to increasing and decreasing BII populations, respectively. The most marked variations are the positive $\Delta\delta P$ s observed on steps with high or very high BI population at 20 °C (δP s typically between -4.5 and -4.2 ppm; see Figure 5). This behavior was observed before, above 20° for BI-rich steps.⁵⁸ Although a full understanding of this observation will ultimately require a detailed analysis, our results show that BI-rich steps gain some BII character under external stress—here, the temperature increase. Yet, one BI-rich step, A₇pA₈ in Oligo 3, shows an exceptional $\Delta\delta P$ of +0.12 ppm (Figure 5 and Figure S1 in Supporting Information). The examination of 1D ¹H spectra led to the detection of slight H2 and H8 resonance broadenings on A₇ and A₈, which may indicate transient excursions toward unknown low-populated conformations requiring sophisticated NMR experiments to be confirmed.⁵⁹ For this reason, the $\Delta\delta P$ corresponding to A₇pA₈ was discarded from the analysis.

The linear fit of the correlation between δP_{20° and $\Delta\delta P$ s is characterized by a correlation coefficient of 0.76 and a standard deviation of 0.10 ppm. This fit enables to estimate that $\Delta\delta P = 0$ (the point for which no change in δP occurs with temperature variation) corresponds to a δP value of -4.01 ± 0.10 ppm at 20 °C (Figure 5). Interestingly, this δP value is very similar to the average δP value previously found for random coil DNAs,

which is centered around 4.0 ppm, with a sequence dependent range of ± 0.2 ppm.⁵⁷ According to eqs 1 and 2, a δP value of -4.01 ± 0.10 ppm corresponds to $BII\% = 48 \pm 7$ (eq 1) or $BII\% = 55 \pm 6$ (eq 2). Thus, both eqs 1 and 2 predicted reasonable $BII\%$, compatible with what is expected for $\Delta\delta P = 0$, i.e., equal populations of BI and BII. This indicates that neglecting a possible sequence effect on δP s of pure BI and pure BII is tenable in practice.

Since the data are not sufficient to discriminate between eqs 1 and 2, in the remainder of the present study, eq 1³⁸ was used to convert to BII percentages the δP s collected at 20 °C.

The average BII percentages ($BII\%_{av}$) were calculated from the average δP s (δP_{av}), themselves deduced from the 72 δP s collected on the four dodecamers at 20, 30, and 40 °C. $BII\%_{av}$ are 16.2 ($\delta P_{av} = -4.23$ ppm) at 20°, 19.4 ($\delta P_{av} = -4.20$ ppm) at 30 °C, and 22.2 ($\delta P_{av} = -4.19$ ppm) at 40 °C. These values are broadly consistent with the $\sim 20\%$ of BII conformers inferred from statistics from X-ray structures⁴⁰ or from δP in solution,^{34,38} without considering the temperature. Superficially, that would initially suggest that there is no distinctive pattern regarding the BI–BII character in sequence 601. Yet, the BII percentages vary considerably along the dodecamer sequences and some phosphate groups exhibit 40% or more of BII conformers (Figure 6). These variations hide a strong sequence effect, which is addressed below.

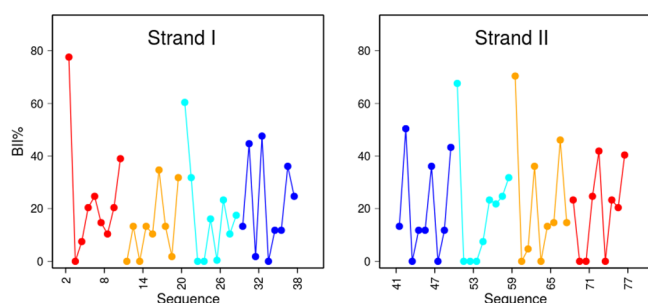


Figure 6. Experimental BII percentages along the four dodecamers. The BII percentages ($BII\%$) were inferred from the ^{31}P chemical shifts at 20 °C and plotted along the two strands of the oligomers, shown in order from end to end (numbering of strand I detailed in Table 1). The terminal dinucleotides, subject to end effects, were not considered. Oligo 1 is in red, Oligo 2 in orange, Oligo 3 in cyan, and Oligo 4 in blue. The left and right panels correspond to the first and the second 5' → 3' strands, respectively.

Validation of the DNA Sequence Effect on the Backbone Behavior with the Dodecamers. Here, we consider the complementary dinucleotides, $Np_iN \bullet Np_jN$, rather than the dinucleotides isolated from their partner. This approach is justified since the B-DNA mechanics imparts a strong coupling between the conformational states of two facing phosphate linkages, $BI \bullet BI$, $BI \bullet BII$, and $BII \bullet BII$; in turn, these combinations of facing backbone states are coupled to the inter base pair parameters of slide, roll, and twist.^{38,40–46} X-ray structure analysis showed that the facing backbone combinations delineate zones in the roll/twist space with negligible overlaps.^{32,40,41} In DNA studied by modeling, $BII \bullet BII$ are rare, but helical parameter values, especially for the twist, can be decomposed into Gaussian-like distributions corresponding to the $BI \bullet BI$ and $BI \bullet BII$ states.⁴⁶ Considering the double helix structure, it is thus more relevant to consider the BII propensities of two facing phosphates than those of each

phosphate separately. In addition, two facing phosphates tend to behave similarly.³⁴ This property is retrieved with the data collected here, the difference between the BII percentages of two facing phosphates being on average 15%. Hence, each $Np_iN \bullet Np_jN$ is characterized by the half-sum of the BII percentages of its facing phosphates i and j ($BII_{ij}\% = (BII_i\% + BII_j\%)/2$).

It has emerged that the $BI \leftrightarrow BII$ equilibrium primarily depends on the dinucleotide sequence.^{34,38,40} From an extensive δP data set, the 10 complementary dinucleotides composing B-DNA were characterized each by specific $BII_{ij}\%$ values, presented as the TRX scale.³⁴ For convenience, these values are reported in Table 2, introducing explicitly the effect of the 3'- and 5'-nearest neighbors on $CpG \bullet CpG$ and $CpA \bullet TpG$. These two steps are particularly sensitive to their nearest neighbors, 5'-Y/3'-R and 5'-R/3'-Y neighbors increasing and decreasing the BII populations, respectively.^{38,60,61} As this previous study³⁴ proposed that BII propensities can be deduced from the sequence, Table 2 and the underpinning training set of data may be seen as providing predictions which may be tested with subsequently obtained BII_{ij} values.

The four dodecamers contain six complementary dinucleotides present in several copies—6 $ApC \bullet GpT$, $ApG \bullet CpT$ and $GpC \bullet GpC$, 5 $CpG \bullet CpG$ and 4 $ApA \bullet TpT$ and $TpA \bullet TpA$ —which enabled confirmation that each of these complementary dinucleotide types is characterized by an average $BII_{ij}\%$ value (Figure S2 in Supporting Information). More extensively, $BII_{ij}\%$ calculated from all the new δP s measured here on the dodecamers were compared with their predicted counterparts from Table 2. These two quantities coincide for all the dinucleotides composing the four dodecamers (Figure 7). Actually, taking into account the effect of the tetrameric environment on $CpG \bullet CpG$ and $CpA \bullet TpG$ enabled us to obtain a remarkable fit between the profiles of newly determined and predicted BII_{ij} propensities along the four dodecamers (Figure 7). So, the δP s measured here on the dodecamers confirmed and reinforced the previous findings on B-DNA sequence-dependent intrinsic backbone behavior. It essentially demonstrates that the BI–BII character of phosphate linkages in B-DNA free in solution is predictable from the sequence. Thus, it provides a strong ground to investigate how the distribution of the dinucleotides with low and high BII propensities is relevant along any DNA sequence owing to the coupling with DNA shape, which is now examined.

Relation between BII Propensities and Minor Groove Width Validated in Solution with NMR. DNA groove widths are measured between two residues on opposite strands, which are closest in space across the groove. These residues are staggered by two or three nucleotides in the 3' direction.⁶² In free DNAs, the groove dimension measured on $Np_iN \bullet Np_jN$ is related to the conformational states of the phosphate groups of the central dinucleotide of interest and of its 3'- and 5'-neighbors, i.e., $Np_{i-1}Np_iNp_{i+1}N \bullet Np_{j-1}Np_jNp_{j+1}N$, mainly because of the cumulative effect of the backbone conformational states on base-pair displacement.⁴⁷ In other words, BI or BII density within a 4 bp segment is coupled to the groove dimensions.⁴⁷ This led to characterization of $Np_iN \bullet Np_jN$ by an extended BII score, $BII_{ij_ext}\%$, which takes into account the neighbors of $Np_iN \bullet Np_jN$, by averaging the individual BII_{ij} percentages of the three complementary dinucleotides in $Np_{i-1}Np_iNp_{i+1}N \bullet Np_{j-1}Np_jNp_{j+1}N$ ($BII_{ij_ext}\% = (BII_{i,j-1}\% + BII_{ij}\% + BII_{i+1,j+1}\%)/3$; see also Materials and Methods).⁴⁷ A main finding of this previous study⁴⁷ concerned the minor groove width. Classical narrow minor grooves are associated

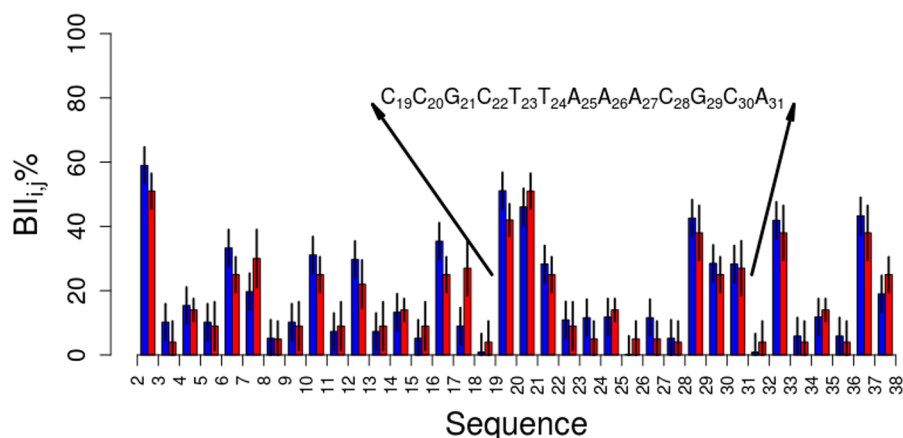
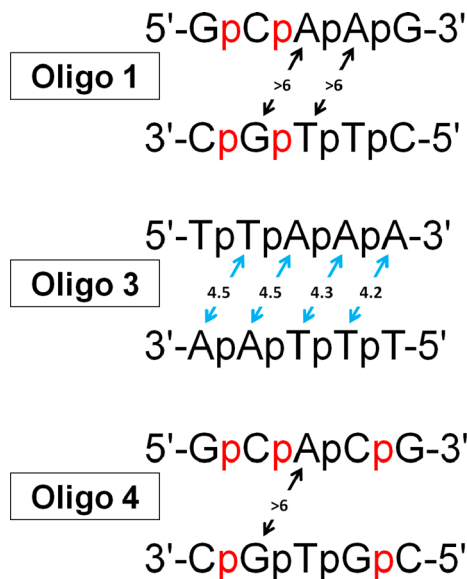


Figure 7. Sequence effect on BII propensities of the complementary dinucleotides forming a part of the 5' half of sequence 601. The complementary dinucleotides are characterized by the half-sum of the BII percentages of their facing phosphates: $BII_{ij}\% = (BII_i\% + BII_j\%)/2$. These quantities are either inferred from the experimental δP values measured in this study (blue bars) or predicted (red bars) from the analysis of an extensive δP data set previously published (Table 2³⁴). The small vertical black bars are the errors, either calculated from the experimental errors on δP s measured in this study for the corresponding $BII_{ij}\%$ or from the standard deviations for predicted $BII_{ij}\%$. The sequence explicitly given is centered on the T:A rich segment located at a 1.5 double-helix turn from the dyad (SHL-1.5) in the NCP. The sequence and the numbering are detailed in Table 1.

with low BII density tetramers (low BII_{ij_ext} values) whereas enlarged minor grooves correspond to segments with high BII density (high BII_{ij_ext} values),⁴⁷ as illustrated in Figure 2.

In NMR, the minor groove width can be investigated on A•T base pairs through the NOE resonances between the H2 of adenine and the H1' located across the strand (Scheme 1),

Scheme 1. H2–H1' Inter-Strand Distances (Å) in Oligos 1, 3, and 4^a



^aThe BI- and BII-rich phosphate linkages are in black and red, respectively. Black arrows indicate the distances potentially observable in the NMR spectra, but not actually discernible. Blue arrows correspond to measured distances.

which are detectable only in the case of narrow minor groove.⁶³ The NMR data collected here along the free BI-rich TTAAA region of Oligo 3 show observable and measurable H2–H1' interstrand distances in spectra recorded with 100, 200, and 400 ms (Scheme 1, Figure 8). In addition, Oligo 1 and 4 contain adenines that provide interesting information about the minor groove width in BII-rich segments. Even using a long mixing

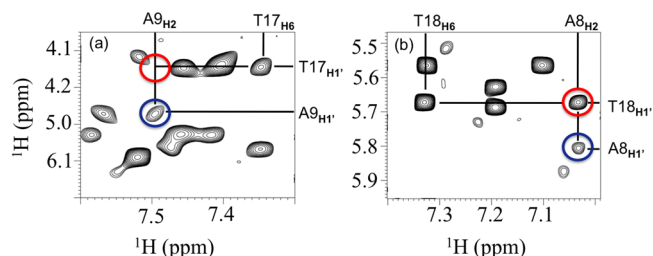


Figure 8. Representative 1H – 1H NOESY spectra of H2–H1' interstrand connectivities. The NOESY spectra show a section of the Base-H1' region obtained for Oligo 1 (left panel) (a) and 3 (right panel) (b) at 20 °C, using a mixing time of 400 ms. In Oligo 1 (left panel) (a), no connectivity is detected between H2 of A9 in strand 1 and H1' of T17 in strand 2. Its theoretical location is indicated by a red circle. By comparison, the intrastrand connectivity observed between H2 and H1' of A9 appears very clearly (blue circle). In Oligo 3 (right panel) (b), the interstrand connectivity observed between H2 of A8 in strand 1 and H1' of T18 in strand 2 (red circle) is stronger than the intrastrand connectivity between H2 and H1' of A8 (blue circle). Such marked cross-peaks also exist for the couples T6–A19, A7–T18, and A9–T16 in Oligo 3. The four interstrand H2–H1' connectivities in Oligo 3 were observed from a mixing time of 100 ms.

time of 400 ms, the H2–H1' interstrand cross-peaks of adenines in or near CpA steps are missing while they could be theoretically observed in the 1H – 1H spectra (Scheme 1, Figure 8). The corresponding distances are thus larger than 6 Å, beyond the limit of NMR detection.

Our measurements thus provide clear evidence that sustained narrow minor grooves exist in free TTAAA elements in solution. They emphasize that a region composed of a succession of BI-rich steps in solution (such as TTAAA) corresponds to a narrow minor groove shape, while BII-rich regions correspond to a wider minor groove. This result validates in solution the relation between BII density and groove shape previously reported on the basis of X-ray structures.⁴⁷

Implications for the Preferential Interaction of DNA Sequences with the Histone Core. We now analyze how the intrinsic structural preferences of free sequences may be relevant for forming the NCP. We focus on the minor groove width because its variation is the most remarkable, recurrent characteristics

of DNA across NCP X-ray structures, regardless of the DNA sequence.^{4–6} This feature is further highlighted by our analysis showing that the correlation coefficient between the minor groove width values calculated on NCPs crystallized with the human α -satellite on one hand and the sequence 601 on the other hand is 0.75. By comparison, the correlation coefficients calculated on the same structures but for roll and slide values are clearly lower, 0.47 and 0.27, respectively. This suggests that the histone core imposes a unique pattern of minor groove variation to DNA upon binding. In NCPs, these conserved wide and narrow minor grooves, which alternate with a periodicity of approximately 10 bp (Figure 9), are located outward and inward from the histone core, respectively.

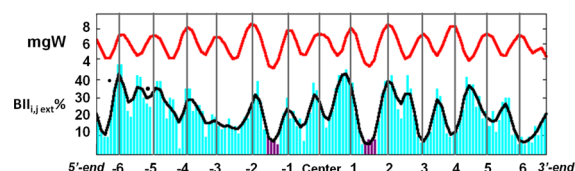


Figure 9. Intrinsic extended BII propensities of the free sequence 601, compared to variation of the minor groove width in the NCP. The intrinsic backbone behavior is expressed in terms of $BII_{ij, ext}\%$ (see the text and Materials and Methods) predicted for each complementary dinucleotide composing the sequence 601, and represented as bars. $BII_{ij, ext}\%$ for the TTAAA elements are in violet, all the others in cyan. Variations in $BII_{ij, ext}\%$ along the sequence are traced with a black line (using a natural smoothing spline approximation to the bars). The sinusoidal variations of the minor groove width (mgW, Å) in the NCP are shown at the top of the panel (red line). These values were calculated and averaged on the NCP X-ray structures containing sequence 601 (PDB entries 3LZ0, 3LZ1, and 3MDV). The vertical gray lines mark the maximal values of minor groove width, expressed in terms of SHL, starting from the center of the sequence 601 (“Center”, at position 73.5 in the 146 bp sequence 601). Apart from the end of the 3’ part of sequence 601 ($SHL > +2.5$), the minor groove distortions in NCP follow the intrinsic extended BII propensities, in turn coupled to groove dimensions in free DNA.

As stated in the previous sections, intrinsic BII propensities can be predicted rather safely from Table 2. The extended BII scores, $BII_{ij, ext}\%$, calculated from the BII% of the dodecamers, match very well their predicted counterparts (Figure S3 in Supporting Information, correlation coefficient of 0.9). $BII_{ij, ext}\%$ prediction was first applied to the whole sequence 601 and compared with the minor groove width profiles extracted from NCP structures crystallized with sequence 601 (Figure 9, Table 4). In this approach, $BII_{ij, ext}\%$ is used as a suitable reporter of the groove shape of free DNA.

In the 5’ part and around the center of sequence 601, most wide or narrow minor grooves in the NCP X-ray structures

Table 4. Correlation Coefficients between the Minor Groove Width of DNA in NCP X-ray Structures and Intrinsic Extended BII Propensities^a

sequence	affinity	whole seq	5’-end	central part	3’-end
High Affinity Sequences					
f2	>601	0.51	0.65	0.73	−0.08
601L	>601	0.42	0.30	0.69	0.30
601	-	0.14	0.25	0.48	−0.63
618	<601	0.10	0.10	0.50	−0.03
Low Affinity Sequences					
h2	≪601	0.07	0.18	0.11	−0.05
h3	≪601	0.16	0.14	0.28	0.12

^aThe minor groove width values were extracted from the NCP X-ray structures containing sequence 601 (PDB entries 3LZ0, 3LZ1, and 3MDV). The intrinsic extended BII propensities, $BII_{ij, ext}\%$ values defined in the text and Materials and Methods, were calculated along artificial sequences displaying either high (601, f2, 601L, 618) or low (h2, h3) affinity for the histone core. The affinities are ranked qualitatively with reference to sequence 601. The correlation coefficients calculated between $BII_{ij, ext}\%$ and minor groove width values were calculated on the whole sequences ($SHL \pm 7$), or only for the 5’-end ($SHL+2.5/SHL+7$), or the central part ($SHL \pm 2.5$), or the 3’ end ($SHL+2.5/SHL+7$).

correspond to peaks of intrinsic high and low BII propensities in free DNA, respectively. The quality of the match is particularly remarkable in the region covering $SHL \pm 2.5$ (Figure 9, Table 4). On the 3’-side of sequence 601, the NCP groove widths at +2.5 helix turns or more from the NCP dyad are clearly out of phase with the corresponding free DNA BII propensities (Figure 9, Table 4). That is not inconsistent with the affinity data, as discussed below.

This analysis suggests that a part of the free sequence 601, in particular, its center, is predisposed to adopt an alternation of wide and narrow minor grooves favorable to NCP formation. To further investigate this idea, five artificial DNA sequences (Table 3) of various affinities (Table 4) for the histone core were examined. Sequence f2 is derived from sequence 601 by an enrichment in TpA steps and mutations in its 5’-part (http://genie.weizmann.ac.il/pubs/nucleosomes06/segal06_data.html). Sequence 601L is the palindromic derivative of the 5’ half of the sequence 601.²⁸ Both f2 and 601L displayed an affinity above that of their parent 601. Sequence 618 has a slightly lower affinity than sequence 601,²⁰ but still a good affinity for NCP formation. In contrast, sequences h2 and h3 have a very low ability to form nucleosomes (http://genie.weizmann.ac.il/pubs/nucleosomes06/segal06_data.html).

Free sequences 601, f2, 601L, and 618 have in common an alternation of minor groove widths in the NCPs, consonant with the variations of intrinsic $BII_{ij, ext}\%$ in the region covering

Table 3. Nucleosomal Sequences^a

Sequence 618	CTGGCGCCTTTTCAAAGTTGTACCTGACCGAGCAGGTGCCTACAGATCCAGACGAGCGGAAATGCCAGAAATTGCGTCTACA-GACCGCTAAGCTCATCTAGAGCTCCCTAGAGCCACCAAGGGCTGTTCCAGAAATTGTCGTAGAA
Sequence 601L	ATCACAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCCTTAACGCACGTACGGAATCCGTACGTGC-GTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGCCCTCGGCACCGGGATTGTGAT
Sequence f2	CTGGAGATACCCGGTGCTAAGGCCGCTTAATTGGTCGTAGCAAGCTCTAGCACCCTTAACGCACGTACGCGCTGTCTACCGC-GTTTAAACCGCAATAGGATTACTTACTAGTCTCTAGGCACGTGAAGATATATACATCCTGT
Sequence h2	CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGATCCTAGCAAGCTCTAGGTGCGCTTAACGCCTGTAGACGCCCTATCCTGTA-CGGCAGTTTAAGCGCACCTAGAGCCTCCGGAATTCACCACGTGTCAGATATATACATCCTGT
Sequence h3	CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGATCCTAGCATACTCTAGGTTAGCTTAACCTACTGTAGACTTACTGTACGGCAG-TTTAAGCTAACCTAGAGTACCCTCTCCGGAATTCACCACGTGTCAGATATATACATCCTGT

^aFive artificial DNA sequences were selected to investigate to what extent their intrinsic properties relate to their affinities for the histone core: sequence 618,²⁰ sequence 601L,²⁸ and sequences f2, h2, and h3 (http://genie.weizmann.ac.il/pubs/nucleosomes06/segal06_data.html).

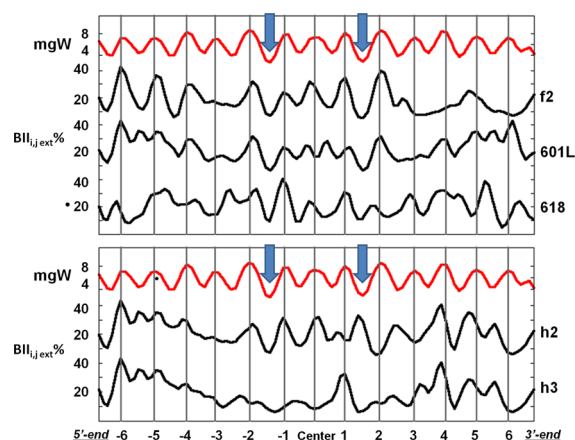


Figure 10. Intrinsic extended BII propensities of selected sequences, compared to variation of the minor groove width in the NCP X-ray structures. The sinusoidal variation of the minor groove width (mgW, Å) in the NCP structures is shown at the top of the panels (red line). The intrinsic backbone behavior is expressed in terms of BII_{ij ext} % (see the text and Materials and Methods); variation along the sequences is traced with black lines (after smoothing with a spline approximation to the bars as shown in Figure 9). The vertical gray lines mark the maximal values of minor groove width, expressed in terms of SHL. The blue arrows indicate the location of SHL ± 1.5. The minor groove shape variations in the NCPs mainly follow the intrinsic extended BII propensities with sequences f2, 601L, and 618, but not sequences h2 and h3. That is consistent with sequences f2, 601L, and 618 being favorable to the NCP formation, while sequences h2 and h3 have a low affinity for the histone core.

SHL ± 2.5 (Figure 10, Table 4). Compared to sequence 601, this match extends on the 5' side with f2 (Figure 10, Table 4). 601L, by duplicating the 5' part of 601, improves the correlation along the 3'-side (Figure 10, Table 4). In contrast with these sequences best at forming NCP, the parallel between intrinsic BII_{ij ext} % and minor groove width in NCPs is considerably weakened with both h2 and h3 (Figure 10, Table 4). A prominent discrepancy concerns SHL+1.5, which is composed of intrinsic BII-rich steps in h2 instead of intrinsic BI-rich steps in high affinity sequences (Figure 10). In h3, SHL+1.5 is composed of intrinsic BI steps, but a marked deficit of intrinsic BII steps characterizes its 3' neighboring segment (Figure 10).

The pictorial analysis presented in Figure 10 was quantified by the regression correlation coefficients between intrinsic BII_{ij ext} % and minor groove width in NCP. For the six considered sequences, the correlation coefficients parallel their affinities for the histone core (Table 4). This strongly supports the notion that the sequence-dependent intrinsic BII propensity of the nucleosomal sequences plays a role in NCP formation via its coupling with minor groove dimension in free DNAs. This explains why the palindromic derivative of the 5' half of the 601 sequence (601L) displays a better affinity than its parent 601, and why the affinity of the palindromic derivative of the 3' half is strongly reduced.²⁸ Also, the recurrence of a maximal agreement in the region SHL ± 2.5 (central 70 bp) in high affinity sequences would account for the remarkable conservation of this region in artificial DNAs selected for their maximal ability to form nucleosomes.²⁰

The above results also support the importance of SHL ± 1.5 sequences. The SHL ± 1.5 regions in sequences 601, 601L, and f2 correspond to two TTAAA fragments. According to the three nucleosome structures crystallized with sequence 601,^{26,27} TTAAA segments robustly anchor histones H3 and H4 owing

to hydrophobic interactions and an electrostatic interaction between Arg63 of H3 and the TTAAA minor groove.^{5,28} The bound TTAAA segments are characterized by sustained narrow minor grooves (3–4 Å) and variable negative rolls (ApA•TpT: $-6.2 \pm 5.8^\circ$; TpA•TpA: $-8.9 \pm 8.8^\circ$) associated with the curvature. On the trinucleotide fragment TAA•TTA, these negative rolls are associated with BI•BII facing phosphate linkages, following the intrinsic trend in free DNA.^{34,38,40,41}

The present NMR data show that the whole free TTAAA element is composed of BI-rich steps (Figure 7) associated with a narrow minor groove (Scheme 1). BI-rich TpA and ApA•TpT dinucleotides intrinsically favor null or positive rolls.^{38,64} The structures of oligomers containing T_nA_n crystallized in various space groups corroborate the presence of positive rolls in such sequences.⁶⁵ So, free and bound TTAAA elements share the narrowing of the minor groove but not the rolls. The deformation of intrinsic null or positive rolls toward negative rolls in the NCPs may have an energetic cost. This cost may be compensated by the negative electrostatic potential generated by the DNA minor grooves, which is especially enhanced in the case of narrow A•T-rich minor grooves.⁶⁶ Such strong electronegative potential associated with the pre-existing narrow minor groove of TTAAA segments could drive the recognition of Arg63 of H3, initializing the anchoring of histones and counterbalancing the cost incurred by negative rolls.

The SHL ± 1.5 elements are surrounded by G•C rich sequences (in sequence 601: CCGCTTAAACGCA in SHL-1.5 and GCGTTTAAACCGC in SHL+1.5) that, in the NCP, adopt the widest minor grooves of ~8 Å (Figures 9 and 10). This is also consistent with the intrinsic structural preferences of free DNA. The flanking regions of SHL-1.5 and SHL+1.5 are composed of dinucleotides with BII as stable as, or even more stable than, BI (Table 2 and Figure 7). Such BII-rich regions in free DNA explore a large conformational landscape that produces minor grooves on average clearly wider than BI-rich segments, as shown in this work. The presence of two or more phosphate linkages simultaneously in BII within a tetrameric segment is in fact associated with minor groove widths reaching 7–8 Å,⁴⁷ as large as those observed in the NCPs.

In sum, both TTAAA elements and their flanking regions have intrinsic minor groove properties that are favorable to nucleosome formation. A key point concerns the two TTAAA motifs, symmetrically located around the nucleosome dyad (SHL ± 1.5) and distant by 30 bp, which allow the anchoring of the two Arg63 of H3 dimer in their narrow, negatively charged, minor groove. Such minor groove intrinsic characteristics are retrieved on another A•T-rich element, GATTA at SHL+3 (Figure 9), which could thus also accommodate one arginine. However, these favorable intrinsic properties are lost in the corresponding 30 bp upstream region, CGCTG at SHL 0, predicted to adopt a wide minor groove with a weakened negative potential. Within the whole sequence 601, TTAAA elements at SHL ± 1.5 emerge as a unique, attractive combination to capture the two Arg63 of H3 dimer.

Taken together, the above results strongly suggest that the sequence-dependent intrinsic structural properties of free DNA contribute to the particularly strong affinity of the sequence 601 and its derivatives for the histone core. They are consistent with a crucial role of pre-existing structural preferences for a regular alternation of wide and narrow minor grooves for NCP indirect readout.

CONCLUSION

Using NMR, we studied four DNA dodecamers relevant to a better understanding of the formation of nucleosome core particles (NCPs). When juxtaposed (discarding the terminal base pairs) these dodecamers recombine a 39 bp fragment of the 5' part of the sequence 601 of very high affinity for the histone core.²⁰ To investigate how DNA intrinsic properties affect the ability to form nucleosomes, one exploited the BI ↔ BII equilibrium of the phosphate groups, which is convenient and reliable reporter of the behavior of dinucleotides in B-DNA.^{34,38,47} The results provide convincing evidence that the sequence-dependent minor groove shape influences the indirect readout mechanism underlying the nucleosome formation.

The studied dodecamers contain dinucleotides with BI-rich (high-field shifted δ Ps) or BII-rich (downfield shifted δ Ps) phosphate linkages. The examination of δ P changes in function of temperature provided additional support of the reliability of the methods used to derive BII populations from δ Ps.^{38,39} Importantly, the newly collected δ Ps enabled us to validate the sequence dependence of the BII propensities. It confirmed that it is dominated by the dinucleotide sequence and therefore is predictable, as previously proposed.³⁴ From a structural point of view, the NMR data validate the proposition based on X-ray structure analyses⁴⁷ that the phosphate groups inform on the minor groove width. As a result, high and low BII densities in DNA tetramer segments free in solution are coupled to wide and narrow minor grooves, respectively.

Taken together, these results offer a specific and powerful framework to address the role of minor groove shape in the nucleosome formation process. The prediction of intrinsic BII propensities was applied to several sequences with various abilities at forming nucleosomes. This approach does not mean that the histone core directly recognizes the backbone states themselves, nor does it suggest that the BII locations should be similar in free and nucleosomal DNAs. Indeed, the backbone angles behave differently in free and bound DNAs, with in particular the emergence of numerous α/γ conformers typical of bound DNAs.^{40,67,68} In our approach, the intrinsic BII propensities are considered reporters of helicoidal features, in particular, the minor groove width.

Only sequences of high affinity for the histone core were found to exhibit intrinsic BII density profiles—representative of minor groove width profiles—that largely parallel the recurrent, characteristic variations of minor groove width in NCP X-ray structures (Figures 9 and 10). The maximal agreement observed along the central ~70 bp is particularly interesting because this region is contacted at the first stage of the nucleosome formation^{23,24} and highly conserved in the highest-affinity artificial sequences.²⁰ Our results also provide new insights on the lexicographic periodicity of ~10 bp found in nucleosomal positioning sequences, *in vitro* and *in vivo*. Such positioning sequences consist of regions enriched in CpG•CpG, GpC•GpC, CpA•TpG, and GpC•CpC that alternate with ApA•TpT, TpA•TpA, and ApT•ApT rich segments.^{9–18} According to our analysis, the first group of dinucleotides imparts free DNA segments with high BII density associated with wide minor grooves; the second group forms BI-rich regions with intrinsic narrow minor grooves. The alternation of both groups seems essential since sequences containing numerous $A_n \bullet T_n$ tracts, unable to spontaneously generate wide minor groove, disfavor nucleosome deposition.^{69,70} Finally, our findings are consistent with a study reporting that the narrowest minor grooves in free

yeast sequences known to be occupied by nucleosomes *in vivo* are compatible with their counterparts in NCP DNA.⁷¹ Hence, the minor groove width in free B-DNA appears as a key structural determinant in the indirect readout process underlying the NCP formation.

Indirect readout exploiting variations of minor groove shape was also demonstrated on the DNA–DNase I interaction.⁵⁰ This enzyme interacts with any DNA minor groove but preferentially cuts sequences adjacent to BII-rich regions, because widening of the minor groove favors the binding of DNase I.⁵⁰ Also, the architectural bacterial protein Fis binds DNA with no obvious sequence selectivity but prefers A•T-rich regions with intrinsic narrow minor groove; three or more A•T → G•C base pair substitutions severely decrease Fis binding by altering the minor groove shape.⁷² The concept of groove shape indirect readout was generalized based on an analysis of exhaustive sets of X-ray structures.^{30,47} Thus, the crucial role of preformed grooves in DNA nucleosomal positioning sequences is not an isolated case. The ability to predict the DNA minor groove shape directly from its sequence not only provides a basis to interpret the effect of mutations on nucleosome formation, but also may help to uncover more examples where this shape is critical to DNA recognition.

In vitro nucleosome binding experiments showed that the temperature affects the nucleosome formation and suggested a role of entropy in this process.⁷³ Also, ions can potentially influence nucleosome positioning.^{2,74} The observations that temperature (this study and ^{39,58}) and cations^{64,75} modulate the phosphate group behavior and consequently the DNA shape are very interesting in this context, although more work is needed to determine the extent to which these factors might influence the nucleosome formation.

The nucleosome organization *in vivo* is likely determined through an interplay between intrinsic DNA properties, chromatin remodelers, and competition with nonhistone proteins bound to DNA.⁷⁶ The relative importance of these factors contributing to the positioning remains controversial. Yet, recent studies confirm that DNA sequence does play a role in positioning nucleosomes *in vivo*, by influencing nucleosome assembly, stability, and disassembly.^{10,76–78} Extensive mappings of nucleosome positioning highlight the heterogeneity of such positioning, with regions of precise and high nucleosome occupancy and other regions where the positioning is much more diffuse.^{79–81} The approach described in the present work could help interpret the large data from such mappings, especially since the proposed sequence-based approach lends itself to high-throughput computational processing.

ASSOCIATED CONTENT

Supporting Information

Tables S1 and S2: ³¹P chemical shifts determined in this study. Figure S1: Variations of internucleotides distances depending on BII percentages. Figure S2: Sequence effect on BII propensities for complementary dinucleotides present in several copies in the studied dodecamers. Figure S3: Sequence effect on extended BII propensities of the complementary dinucleotides of the four dodecamers. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: olivier.mauffret@lbp.ens-cachan.fr. Tel.: +33-147407421. Fax: +33-147407671.

*E-mail: bhartman@ens-cachan.fr. Tel.: +33-147407421. Fax: +33-147407671.

Author Contributions

Xiaoqian Xu and Akli Ben Imeddourene have equally contributed to the work. Akli Ben Imeddourene and Xiaoqian Xu performed the NMR experiments under the supervision of Loussiné Zargarian and Olivier Mauffret. Akli Ben Imeddourene carried out most analyses. Brigitte Hartmann, the project leader, wrote the manuscript with Nicolas Foloppe.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The East China Normal University is gratefully acknowledged for its support. The authors thank Drs Sylvie Rimsky (LBPA, CNRS/ENS de Cachan) and Christophe Oguey (LPTM, Université de Cergy-Pontoise) for helpful advice.

REFERENCES

- (1) Andrews, A. J., and Luger, K. (2011) Nucleosome structure(s) and stability: variations on a theme. *Annu. Rev. Biophys.* 40, 99–117.
- (2) Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W., and Richmond, T. J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* 319, 1097–1113.
- (3) Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- (4) Olson, W. K., and Zhurkin, V. B. (2011) Working the kinks out of nucleosomal DNA. *Curr. Opin. Struct. Biol.* 21, 348–357.
- (5) Wu, B., Mohideen, K., Vasudevan, D., and Davey, C. A. (2010) Structural insight into the sequence dependence of nucleosome positioning. *Structure* 18, 528–536.
- (6) Xu, F., and Olson, W. K. (2010) DNA architecture, deformability, and nucleosome positioning. *J. Biomol. Struct. Dyn.* 27, 725–739.
- (7) Balasubramanian, S., Xu, F., and Olson, W. K. (2009) DNA sequence-directed organization of chromatin: structure-based computational analysis of nucleosome-binding sequences. *Biophys. J.* 96, 2245–2260.
- (8) Richmond, T. J., and Davey, C. A. (2003) The structure of DNA in the nucleosome core. *Nature* 423, 145–150.
- (9) Widom, J. (2001) Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* 34, 269–324.
- (10) Brogaard, K., Xi, L., Wang, J. P., and Widom, J. (2012) A map of nucleosome positions in yeast at base-pair resolution. *Nature* 486, 496–501.
- (11) Cui, F., and Zhurkin, V. B. (2010) Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro. *J. Biomol. Struct. Dyn.* 27, 821–841.
- (12) Drew, H. R., and Calladine, C. R. (1987) Sequence-specific positioning of core histones on an 860 base-pair DNA. Experiment and theory. *J. Mol. Biol.* 195, 143–173.
- (13) Drew, H. R., and Travers, A. A. (1985) DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* 186, 773–790.
- (14) Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., LeProust, E. M., Hughes, T. R., Lieb, J. D., Widom, J., and Segal, E. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458, 362–366.
- (15) Satchwell, S. C., Drew, H. R., and Travers, A. A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* 191, 659–675.
- (16) Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I. K., Wang, J. P., and Widom, J. (2006) A genomic code for nucleosome positioning. *Nature* 442, 772–778.

- (17) Takasuka, T. E., and Stein, A. (2010) Direct measurements of the nucleosome-forming preferences of periodic DNA motifs challenge established models. *Nucleic Acids Res.* 38, 5672–5680.
- (18) Travers, A., Hiriart, E., Churcher, M., Caserta, M., and Di Mauro, E. (2010) The DNA sequence-dependence of nucleosome positioning in vivo and in vitro. *J. Biomol. Struct. Dyn.* 27, 713–724.
- (19) Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., and Zhurkin, V. B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.* 95, 11163–11168.
- (20) Thastrom, A., Bingham, L. M., and Widom, J. (2004) Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. *J. Mol. Biol.* 338, 695–709.
- (21) Miele, V., Vaillant, C., d'Aubenton-Carafa, Y., Thermes, C., and Grange, T. (2008) DNA physical properties determine nucleosome occupancy from yeast to fly. *Nucleic Acids Res.* 36, 3746–3756.
- (22) Hall, M. A., Shundrovsky, A., Bai, L., Fulbright, R. M., Lis, J. T., and Wang, M. D. (2009) High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat. Struct. Mol. Biol.* 16, 124–129.
- (23) Cremisi, C., and Yaniv, M. (1980) Sequential assembly of newly synthesized histones on replicating SV40 DNA. *Biochem. Biophys. Res. Commun.* 92, 1117–1123.
- (24) Worcel, A., Han, S., and Wong, M. L. (1978) Assembly of newly replicated chromatin. *Cell* 15, 969–977.
- (25) Fernandez, A. G., and Anderson, J. N. (2007) Nucleosome positioning determinants. *J. Mol. Biol.* 371, 649–668.
- (26) Makde, R. D., England, J. R., Yennawar, H. P., and Tan, S. (2010) Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* 467, 562–566.
- (27) Vasudevan, D., Chua, E. Y., and Davey, C. A. (2010) Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol.* 403, 1–10.
- (28) Chua, E. Y., Vasudevan, D., Davey, G. E., Wu, B., and Davey, C. A. (2012) The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res.* 40, 6338–6352.
- (29) Ong, M. S., Richmond, T. J., and Davey, C. A. (2007) DNA stretching and extreme kinking in the nucleosome core. *J. Mol. Biol.* 368, 1067–1074.
- (30) Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., and Mann, R. S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* 79, 233–269.
- (31) West, S. M., Rohs, R., Mann, R. S., and Honig, B. (2010) Electrostatic interactions between arginines and the minor groove in the nucleosome. *J. Biomol. Struct. Dyn.* 27, 861–866.
- (32) Foloppe, N., Gueroult, M., and Hartmann, B. (2013) Simulating DNA by molecular dynamics: aims, methods, and validation. *Methods Mol. Biol.* 924, 445–468.
- (33) Heddi, B., Foloppe, N., Oguey, C., and Hartmann, B. (2008) Importance of accurate DNA structures in solution: the Jun-Fos model. *J. Mol. Biol.* 382, 956–970.
- (34) Heddi, B., Oguey, C., Lavelle, C., Foloppe, N., and Hartmann, B. (2010) Intrinsic flexibility of B-DNA: the experimental TRX scale. *Nucleic Acids Res.* 38, 1034–1047.
- (35) Fratini, A. V., Kopka, M. L., Drew, H. R., and Dickerson, R. E. (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.* 257, 14686–14707.
- (36) Gorenstein, D. G. (1984) *Phosphorus-31 NMR: Principles and Applications*, Academic Press, New York.
- (37) Gorenstein, D. G. (1992) ³¹P NMR of DNA. *Methods Enzymol.* 211, 254–286.
- (38) Heddi, B., Foloppe, N., Bouchemal, N., Hantz, E., and Hartmann, B. (2006) Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition. *J. Am. Chem. Soc.* 128, 9170–9177.
- (39) Tian, Y., Kayatta, M., Shultis, K., Gonzalez, A., Mueller, L. J., and Hatcher, M. E. (2009) ³¹P NMR investigation of backbone dynamics in DNA binding sites†. *J. Phys. Chem. B* 113, 2596–2603.

- (40) Djuranovic, D., and Hartmann, B. (2003) Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *J. Biomol. Struct. Dyn.* 20, 771–788.
- (41) Djuranovic, D., and Hartmann, B. (2004) DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers* 73, 356–368.
- (42) Hartmann, B., Piazzola, D., and Lavery, R. (1993) BI-BII transitions in B-DNA. *Nucleic Acids Res.* 21, 561–568.
- (43) Srinivasan, A. R., and Olson, W. K. (1987) Nucleic acid model building: the multiple backbone solutions associated with a given base morphology. *J. Biomol. Struct. Dyn.* 4, 895–938.
- (44) van Dam, L., and Levitt, M. H. (2000) BII nucleotides in the B and C forms of natural-sequence polymeric DNA: A new model for the C form of DNA. *J. Mol. Biol.* 304, 541–561.
- (45) Winger, R. H., Liedl, K. R., Pichler, A., Hallbrucker, A., and Mayer, E. (1999) Helix morphology changes in B-DNA induced by spontaneous B(I) \rightleftharpoons B(II) substrate interconversion. *J. Biomol. Struct. Dyn.* 17, 223–235.
- (46) Drsata, T., Perez, A., Orozco, M., Morozov, A. V., Sponer, J., and Lankas, F. (2013) Structure, stiffness and substates of the dickerson-drew dodecamer. *J. Chem. Theory Comput.* 9, 707–721.
- (47) Oguey, C., Foloppe, N., and Hartmann, B. (2010) Understanding the sequence-dependence of DNA groove dimensions: implications for DNA interactions. *PLoS One* 5, e15931.
- (48) Prive, G. G., Yanagi, K., and Dickerson, R. E. (1991) Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-T-T-G-G. *J. Mol. Biol.* 217, 177–199.
- (49) Wang, D., Ulyanov, N. B., and Zhurkin, V. B. (2010) Sequence-dependent Kink-and-Slide deformations of nucleosomal DNA facilitated by histone arginines bound in the minor groove. *J. Biomol. Struct. Dyn.* 27, 843–859.
- (50) Heddi, B., Abi-Ghanem, J., Lavigne, M., and Hartmann, B. (2010) Sequence-dependent DNA flexibility mediates DNase I cleavage. *J. Mol. Biol.* 395, 123–133.
- (51) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293.
- (52) Piatto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2, 661–665.
- (53) Baleja, J. D., Moulton, J., and Sykes, B. D. (1990) Distance measurement and structure refinement with Noe data. *J. Magn. Reson.* 87, 375–384.
- (54) Sklenar, V., Miyashiro, H., Zon, G., Miles, H. T., and Bax, A. (1986) Assignment of the 31P and 1H resonances in oligonucleotides by two-dimensional NMR spectroscopy. *FEBS Lett.* 208, 94–98.
- (55) Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D., and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* 37, S917–S929.
- (56) Precechtelova, J., Munzarova, M. L., Vaara, J., Novotny, J., Dracinski, M., and Sklenar, V. (2013) Toward reproducing sequence trends in phosphorus chemical shifts for nucleic acids by MD/DFT calculations. *J. Chem. Theory Comput.* 9, 1641–1656.
- (57) Ho, C. N., and Lam, S. L. (2004) Random coil phosphorus chemical shift of deoxyribonucleic acids. *J. Magn. Reson.* 171, 193–200.
- (58) Nikonowicz, E. P., and Gorenstein, D. G. (1990) Two-dimensional 1H and 31P NMR spectra and restrained molecular dynamics structure of a mismatched GA decamer oligodeoxyribonucleotide duplex. *Biochemistry* 29, 8845–8858.
- (59) Nikolova, E. N., Kim, E., Wise, A. A., O'Brien, P. J., Andricioaei, I., and Al-Hashimi, H. M. (2011) Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* 470, 498–502.
- (60) Lefebvre, A., Mauffret, O., Hartmann, B., Lescot, E., and Femandjian, S. (1995) Structural behavior of the CpG step in two related oligonucleotides reflects its malleability in solution. *Biochemistry* 34, 12019–12028.
- (61) Tisne, C., Delepierre, M., and Hartmann, B. (1999) How NF-kappaB can be attracted by its cognate DNA. *J. Mol. Biol.* 293, 139–150.
- (62) Stofer, E., and Lavery, R. (1994) Measuring the geometry of DNA grooves. *Biopolymers* 34, 337–346.
- (63) Chuprina, V. P., Lipanov, A. A., Fedoroff, O., Kim, S. G., Kintanar, A., and Reid, B. R. (1991) Sequence effects on local DNA topology. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9087–9091.
- (64) Heddi, B., Foloppe, N., Hantz, E., and Hartmann, B. (2007) The DNA structure responds differently to physiological concentrations of K(+) or Na(+). *J. Mol. Biol.* 368, 1403–1411.
- (65) Mack, D. R., Chiu, T. K., and Dickerson, R. E. (2001) Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: a comparative analysis of T-A and A-T steps within A-tracts. *J. Mol. Biol.* 312, 1037–1049.
- (66) Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. (2009) The role of DNA shape in protein-DNA recognition. *Nature* 461, 1248–1253.
- (67) Djuranovic, D., and Hartmann, B. (2005) Molecular dynamics studies on free and bound targets of the bovine papillomavirus type I e2 protein: the protein binding effect on DNA and the recognition mechanism. *Biophys. J.* 89, 2542–2551.
- (68) Varnai, P., Djuranovic, D., Lavery, R., and Hartmann, B. (2002) Alpha/gamma transitions in the B-DNA backbone. *Nucleic Acids Res.* 30, 5398–5406.
- (69) Hughes, A., and Rando, O. J. (2009) Chromatin 'programming' by sequence—is there more to the nucleosome code than %GC? *J. Biol.* 8, 96.
- (70) Segal, E., and Widom, J. (2009) Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* 19, 65–71.
- (71) Bishop, E. P., Rohs, R., Parker, S. C., West, S. M., Liu, P., Mann, R. S., Honig, B., and Tullius, T. D. (2011) A map of minor groove shape and electrostatic potential from hydroxyl radical cleavage patterns of DNA. *ACS Chem. Biol.* 6, 1314–1320.
- (72) Stella, S., Cascio, D., and Johnson, R. C. (2010) The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes Dev.* 24, 814–826.
- (73) Wu, C., and Travers, A. (2005) Relative affinities of DNA sequences for the histone octamer depend strongly upon both the temperature and octamer concentration. *Biochemistry* 44, 14329–14334.
- (74) Davey, C. A., and Richmond, T. J. (2002) DNA-dependent divalent cation binding in the nucleosome core particle. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11169–11174.
- (75) Guerault, M., Boittin, O., Mauffret, O., Etchebest, C., and Hartmann, B. (2012) Mg2+ in the major groove modulates B-DNA structure and dynamics. *PLoS One* 7, e41704.
- (76) Struhl, K., and Segal, E. (2013) Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* 20, 267–273.
- (77) Parmar, J. J., Marko, J. F., and Padinhateeri, R. (2014) Nucleosome positioning and kinetics near transcription-start-site barriers are controlled by interplay between active remodeling and DNA sequence. *Nucleic Acids Res.* 42, 128–136.
- (78) Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., and Segal, E. (2012) Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat. Genetics* 44, 743–750.
- (79) Jansen, A., and Verstrepen, K. J. (2011) Nucleosome positioning in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* 75, 301–320.
- (80) Korber, P. (2012) Active nucleosome positioning beyond intrinsic biophysics is revealed by in vitro reconstitution. *Biochem. Soc. Trans.* 40, 377–382.
- (81) Zhang, Z., and Pugh, B. F. (2011) High-resolution genome-wide mapping of the primary structure of chromatin. *Cell* 144, 175–186.

NMR studies of DNA support the role of pre-existing minor groove variations in nucleosome indirect readout

X. Xu, A. Ben Imeddourene, L. Zargarian, N. Foloppe, O. Mauffret, and B. Hartmann

Supporting Information

Tables S1 and S2: ^{31}P chemical shifts determined in this study.

Figure S1: Changes in ^{31}P chemical shifts in response to changes in temperature.

Figure S2: Sequence effect on BII propensities for complementary dinucleotides present in several copies in the studied dodecamers.

Figure S3: Sequence effect on extended BII propensities of the 37 complementary dinucleotides of the four dodecamers.

**Table S1: ^{31}P chemical shifts determined in this study
for Oligo 1 and Oligo 2**

	step	$\delta\text{P}_{20^\circ}$	$\delta\text{P}_{30^\circ}$	$\delta\text{P}_{40^\circ}$		step	$\delta\text{P}_{20^\circ}$	$\delta\text{P}_{30^\circ}$	$\delta\text{P}_{40^\circ}$
Oligo 1, S1	5'-T1pC2	-4.24	-4.17	-4.11	Oligo 2, S1	5'-G1pC2	-4.19	-4.25	-4.21
	C2pG3	-3.80	-3.78	-3.80		C2pT3	-4.45	-4.41	-4.38
	G3pT4	-4.45	-4.42	-4.38		T3pC4	-4.25	-4.22	-4.21
	T4pA5	-4.29	-4.27	-4.24		C4pT5	-4.37	-4.34	-4.32
	A5pG6	-4.20	-4.16	-4.13		T5pA6	-4.25	-4.23	-4.23
	G6pC7	-4.17	-4.15	-4.16		A6pG7	-4.27	-4.24	-4.23
	C7pA8	-4.24	-4.22	-4.21		G7pC8	-4.1	-4.07	-4.07
	A8pA9	-4.27	-4.24	-4.22		C8pA9	-4.25	-4.23	-4.23
	A9pG10	-4.20	-4.16	-4.13		A9pC10	-4.33	-4.3	-4.29
	G10pC11	-4.07	-4.04	-4.02		C10pC11	-4.12	-4.09	-4.08
Oligo 1, S2	C11pT12-3'	-4.18	-4.11	-4.06	Oligo 2, S2	C11pG12-3'	-3.83	-3.91	-3.88
	5'-A13pG14	-4.20	-4.15	-4.12		5'-C13pG14	-4.07	-4.11	-4.09
	G14pC15	-4.18	-4.16	-4.16		G14pG15	-3.85	-3.85	-3.87
	C15pT16	-4.39	-4.35	-4.32		G15pT16	-4.45	-4.42	-4.4
	T16pT17	-4.46	-4.42	-4.39		T16pG17	-4.31	-4.29	-4.27
	T17pG18	-4.17	-4.15	-4.16		G17pC18	-4.09	-4.09	-4.09
	G18pC19	-4.05	-4.02	-4.03		C18pT19	-4.50	-4.46	-4.43
	C19pT20	-4.50	-4.46	-4.42		T19pA20	-4.25	-4.24	-4.23
	T20pA21	-4.18	-4.16	-4.16		A20pG21	-4.24	-4.23	-4.22
	A21pC22	-4.20	-4.16	-4.13		G21pA22	-4.02	-3.99	-3.99
Oligo 1, S2	C22pG23	-4.06	-4.04	-4.03		A22pG23	-4.24	-4.23	-4.22
	G23pA24-3'	-4.11	-4.07	-4.04		G23pC24-3'	-4.05	-4.159	-4.16

The ^{31}P chemical shifts (δP , ppm) were collected at three temperatures (20 °C, 30 °C and 40 °C). δP are referenced to trimethylphosphate. S1 and S2 indicate strand 1 and strand 2, respectively. The sequences of the dodecamers Oligo 1, 2, 3 and 4 overlap by three base pairs, equivalent to two base steps (shaded background).

**Table S2: ^{31}P chemical shifts determined in this study
for Oligo 3 and Oligo 4**

	step	$\delta\text{P}_{20^\circ}$	$\delta\text{P}_{30^\circ}$	$\delta\text{P}_{40^\circ}$		step	$\delta\text{P}_{20^\circ}$	$\delta\text{P}_{30^\circ}$	$\delta\text{P}_{40^\circ}$
Oligo 3, S1	5'-C1pC2	-4.11	-4.08	-4.04	Oligo 4, S1	5'-C1pG2	-4.14	-4.1	-4.05
	C2pG3	-3.92	-3.91	-3.9		G2pC3	-4.25	-4.23	-4.21
	G3pC4	-4.12	-4.11	-4.09		C3pA4	-4.03	-4.03	-4.02
	C4pT5	-4.35	-4.33	-4.3		A4pC5	-4.33	-4.31	-4.27
	T5pT6	-4.46	-4.43	-4.39		C5pG6	-4.01	-3.99	-3.98
	T6pA7	-4.23	-4.23	-4.21		G6pT7	-4.47	-4.44	-4.4
	A7pA8	-4.34	-4.29	-4.22		T7pA8	-4.26	-4.24	-4.22
	A8pA9	-4.18	-4.17	-4.16		A8pC9	-4.26	-4.24	-4.21
	A9pC10	-4.27	-4.23	-4.2		C9pG10	-4.09	-4.07	-4.05
	C10pG11	-4.22	-4.18	-4.15		G10pC11	-4.17	-4.14	-4.11
	G11pC12-3'	-4.07	-4.05	-4.03		C11pG12-3'	-3.98	-3.96	-3.95
Oligo 3, S2	5'-G13pC14	-4.29	-4.25	-4.21	Oligo 4, S2	5'-C13pG14	-4.14	-4.1	-4.05
	C14pG15	-3.87	-3.91	-3.9		G14pC15	-4.25	-4.23	-4.21
	G15pT16	-4.41	-4.38	-4.34		C15pG16	-3.99	-3.98	-3.98
	T16pT17	-4.51	-4.47	-4.42		G16pT17	-4.43	-4.40	-4.36
	T17pT18	-4.41	-4.38	-4.34		T17pA18	-4.26	-4.24	-4.22
	T18pA19	-4.29	-4.26	-4.23		A18pC19	-4.26	-4.24	-4.21
	A19pA20	-4.18	-4.17	-4.16		C19pG20	-4.09	-4.07	-4.04
	A20pG21	-4.19	-4.17	-4.15		G20pT21	-4.47	-4.44	-4.4
	G21pC22	-4.17	-4.15	-4.12		T21pG22	-4.26	-4.24	-4.22
	C22pG23	-4.12	-4.11	-4.09		G22pC23	-4.04	-4.02	-4.00
	G23pG24-3'	-3.93	-3.92	-3.91		C23pG24-3'	-3.87	-3.91	-3.9

The ^{31}P chemical shifts (δP , ppm) were collected at three temperatures (20 °C, 30 °C and 40 °C). δP are referenced to trimethylphosphate. S1 and S2 indicate strand 1 and strand 2, respectively. The sequences of the dodecamers Oligo 1, 2, 3 and 4 overlap by three base pairs, equivalent to two base steps (shaded background).

Figure S1. Changes in ^{31}P chemical shifts in response to changes in temperature. Typical variations of ^{31}P chemical shifts (δP , ppm) as a function of temperature ($^{\circ}\text{C}$) are shown for down-field (GpG, CpA and GpC, $\delta\text{P} > -4.1\text{ppm}$) and high-field (TpA, ApC, CpT, $\delta\text{P} < -4.1\text{ppm}$) shifted phosphate linkages. The phosphate linkage of A_7pA_8 (blue square) in the TTAA segment of Oligo 3 shows an exceptional variation of $+0.12\text{ppm}$ between 20 and 40°C that echoes slight H2 and H8 resonance broadenings observed in $1\text{D } ^1\text{H}$ spectra.

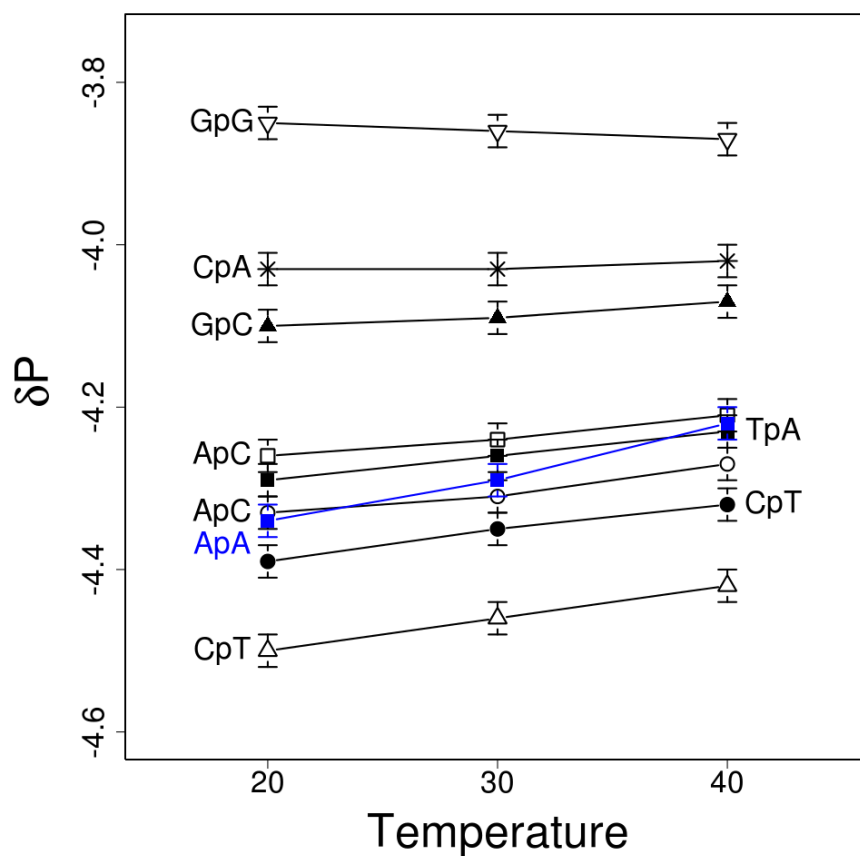


Figure S2: Sequence effect on BII propensities for complementary dinucleotides present in several copies in the studied dodecamers. The complementary dinucleotides $Np_iN \bullet Np_jN$ are characterized by the half-sum of the BII percentages of their facing phosphates: $BII_{i,j}\% = (BII_i\% + BII_j\%)/2$. These quantities are either inferred from the experimental δP values measured in this study (blue points) or predicted (red squares) from the analysis of an extensive δP dataset previously published (Table 2, (34)). The vertical bars are standard deviations.

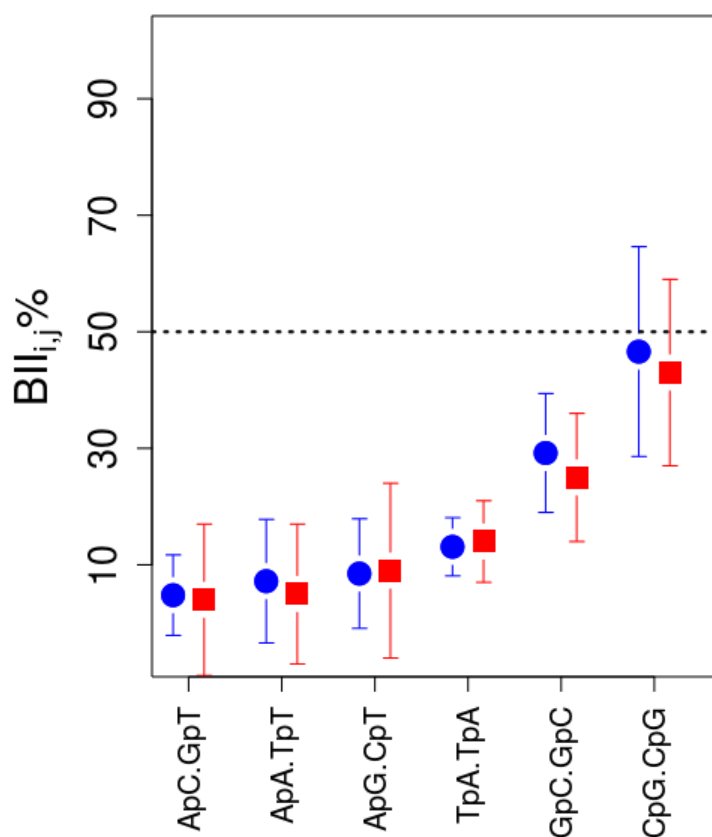
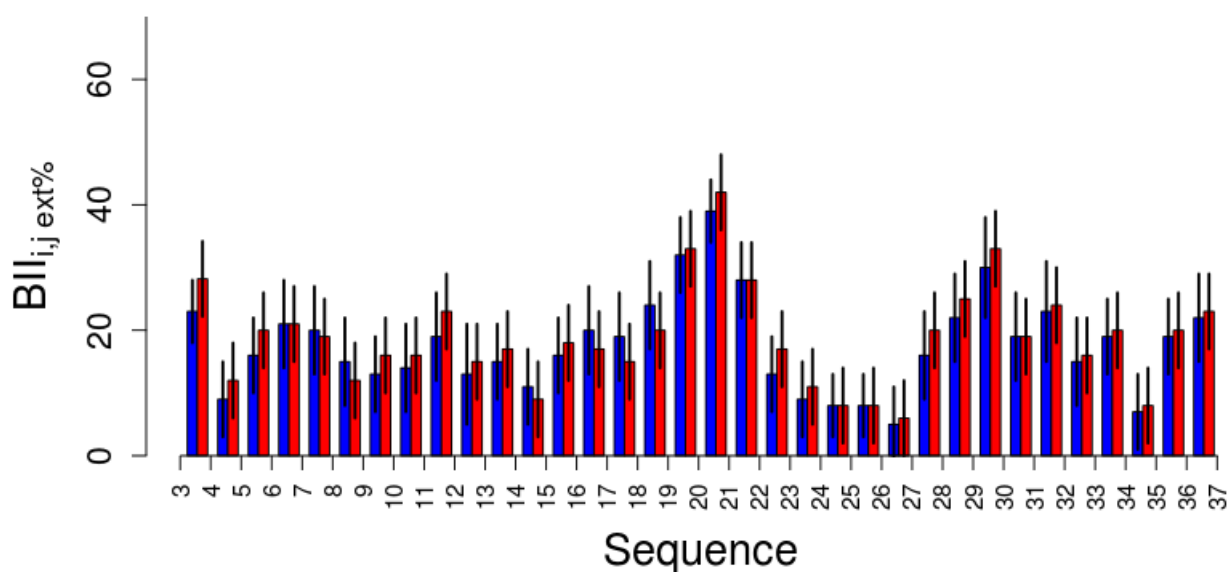


Figure S3: Sequence effect on extended BII propensities of the complementary dinucleotides of the four dodecamers. The complementary dinucleotides $Np_iN \bullet Np_jN$ are characterized by $BII_{i,j \text{ ext}}\%$, the averaged sum of the individual $BII_{i,j}\%$ of the three complementary dinucleotides in the tetramer $Np_{i-1}Np_iNp_{i+1}N \bullet Np_{j-1}Np_jNp_{j+1}N$: $BII_{i,j \text{ ext}}\% = (BII_{i-1,j-1}\% + BII_{i,j}\% + BII_{i+1,j+1}\%) / 3$. These quantities are either inferred from the experimental δP values measured in this study (blue bars) or predicted (red bars) from the analysis of an extensive δP dataset previously published (Table 2, (34)). The small vertical black bars are the errors, either calculated from the experimental errors on δP s measured in this study for the corresponding $BII_{i,j \text{ ext}}\%$ or from the standard deviations for predicted $BII_{i,j \text{ ext}}\%$. The sequence and the numbering are detailed in Table 1.



CHAPITRE 2

Test de champs de force et nouveaux aspects de la dynamique des groupements phosphate.

1 Introduction

La paramétrisation de champs de force empiriques visant à simuler l'ADN en solution a connu une grande expansion ces dernières années. Après avoir pallié à des problèmes liés à la représentation des angles de brin α/γ , l'objectif a été d'augmenter les populations de conformères BII et de reproduire au mieux l'effet de séquence sur les échanges conformationnels BI \leftrightarrow BII. Les champs de force appartenant aux deux grandes familles CHARMM et Parm ont ainsi été récemment recalibrés et deux nouvelles versions sont maintenant disponibles, CHARMM36 (Hart et al. 2012) et Parmbsc0_{ezOLI} (Zgarbová et al. 2013). En plus des comparaisons entre les énergies libres issues de calculs quantiques *ab initio* ou bien calculées à partir des dynamiques moléculaires, les auteurs ont pris le soin de comparer les ratio BI/BII simulés à leurs équivalents expérimentaux. Toutefois, ces tests se focalisent essentiellement sur le dodécamère de Drew-Dickerson qui, du fait de sa notoriété historique, est la séquence la plus étudiée par cristallographie et RMN. Or cette séquence CGCGAATTCGCG est palindromique et ne contient que 7 types de dinucléotides sur les 16 possibles. L'échantillon des pas dinucléotidiques est donc loin d'être complet, et, *a fortiori*, il n'est pas possible d'y étudier un possible effet de l'environnement tétramérique. Pour juger de la qualité de ces nouveaux champs de force, plus de séquences doivent être simulées et testées.

Les déplacements chimiques du phosphore collectés sur nos quatre dodécamères offrent l'opportunité de tester exhaustivement CHARMM36 et Parmbsc0_{ezOLI}. J'ai donc réalisé des simulations de dynamique moléculaire en présence d'eau et de contre-ions avec ces nouveaux champs de force, d'une durée totale de 4.7 μ s.

La comparaison entre les populations de BII simulées et expérimentales a montré un net progrès

par rapport aux champs de force initiaux dont ils sont issus. La principale amélioration consiste en une bonne représentation du pourcentage de BII moyen (calculé sur l'ensemble des dinucléotides), alors qu'avant ce pourcentage était nettement inférieur à celui qu'on obtient soit par RMN soit en analysant les structures cristallographiques. Cependant, l'effet de la séquence n'est pas complètement reproduit, chacun des champs de force ayant ses propres biais.

Toutefois, l'accroissement des populations BII, notamment par CHARMM36, nous a permis d'analyser le comportement dynamique des phosphates face à face (NpN.NpN), une information qui échappe à la RMN. Nous avons ainsi montré par une analyse statistique que deux phosphates qui se font face ont des mouvements décorrélés. En d'autres termes, ces phosphates ont des comportements dynamiques qui sont indépendants l'un de l'autre. Cette observation a pour conséquence qu'il est maintenant possible de calculer très simplement les populations des trois états possibles, BI.BI, BI.BII et BII.BII, à partir des données expérimentales. L'intérêt de cette approche réside dans le fait que, dans les structures cristallographiques, les états conformationnels des phosphates face à face sont couplés à des paramètres hélicoïdaux (*Twist*, *Roll* et *Slide*) (Djuranovic et Hartmann 2003) (Djuranovic et Hartmann 2004). Pour parfaire le test des champs de force, nous avons analysé les dynamiques moléculaires et montré que ces couplages sont reproduits correctement et d'une manière cohérente par CHARMM36 et Parmbsc0_{εζOLI}.

Les dynamiques présentées dans l'article qui suit cette introduction ont toutes été réalisées en maintenant contraints les appariements Watson-Crick de paires de bases terminales. Nous allons expliquer pourquoi.

Avec CHARMM27, les paires de bases terminales et parfois même les paires de bases voisines s'ouvrent dès le début de la trajectoire (Heddi et al. 2008). CHARMM36 montre la même prédisposition aux ouvertures de paires de bases dans tous les essais réalisés avec des points de départ différents. Cependant, avec Parmbsc0_{εζOLI}, les ouvertures des bases terminales n'étaient pas systématiques notamment au début de la dynamique. Mais dans une première série de dynamiques sans aucune contrainte sur les appariements de paires de bases terminales, nous avons vu que les paires de bases s'ouvraient après 100 ou 300 ns (Figure 25).

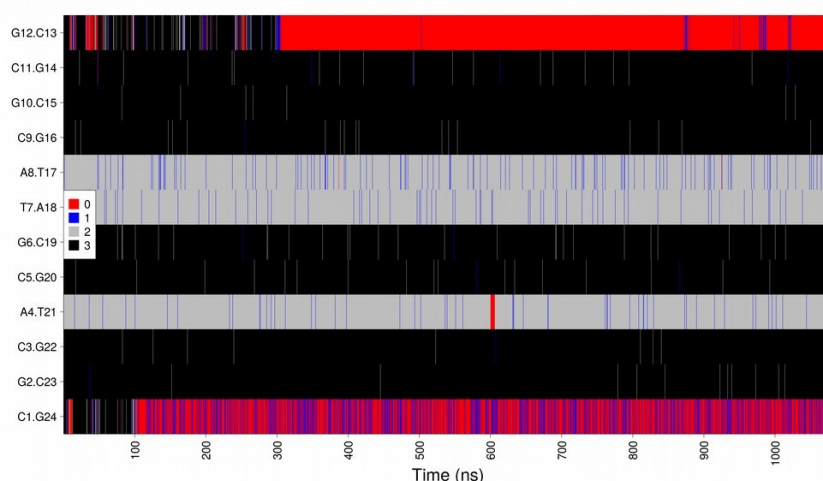


Figure 25: Evolution des liaisons hydrogènes des appariements Watson-Crick de l'ensemble de paires bases d'un des dodécamères (oligo 4) en fonction du temps, au cours d'une microseconde de dynamique moléculaire avec le champ de force Parmbsc0_{εζOLI}, sans contrainte sur les appariements des bases terminales. La légende au centre à gauche montre les couleurs attribuées au nombre de liaisons hydrogènes existantes.

En fait, ces ouvertures s'accompagnent d'importants problèmes de convergence sur un certain nombre de dinucléotides non terminaux (voir les exemples Figure 26), qui, même après une microseconde de dynamique, n'atteignent pas un état d'équilibre. Cette absence de convergence peut être considérée comme anormale, puisque les mouvements du phosphate se produisent à l'échelle de la pico-nanoseconde et qu'une microseconde de trajectoire est théoriquement largement suffisante pour assurer une bonne représentation statistique de ces mouvements.

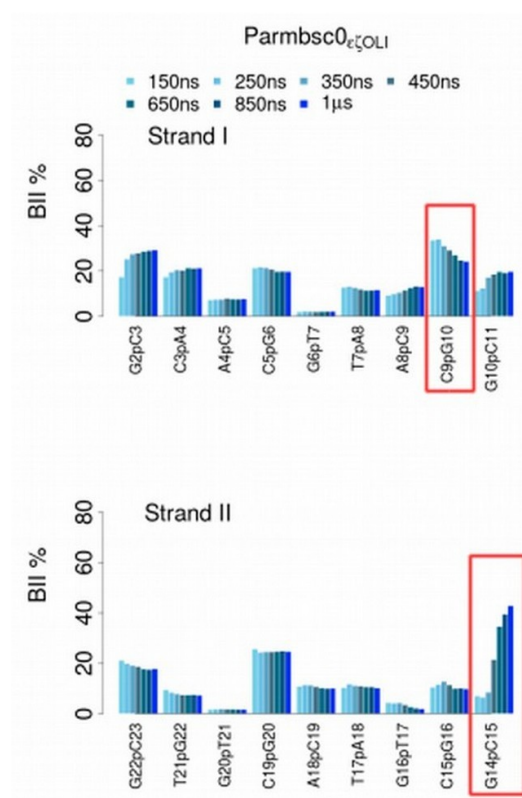


Figure 26 : Evolution des pourcentages de BII dans un des dodécamères (oligo 4) au cours d'une microseconde de dynamique moléculaire avec le champ de force Parmbsc0_{ezOLI}, sans contrainte sur les appariements des bases terminales. Les premiers et derniers pas dinucléotidiques ne sont pas considérés. Les 50 premières nanosecondes correspondant à la phase dite d'équilibration sont éliminées de l'analyse. Les pourcentages de BII sont calculés pour des parties de plus en plus longues de la trajectoire, de 50 à 150ns, de 50 à 250ns, de 50 à 350ns, etc. Lorsque la convergence est bonne, ces pourcentages sont constants. Nous remarquons que dans deux cas (encadrés en rouge), les pourcentages de BII ne sont pas stabilisés.

Nous avons alors observé que les bases désappariées se retournent vers la partie interne de la double hélice pour former des interactions avec des phosphates, s'immiscer dans les sillons ou encore se mésappairer avec des bases. Lors de la trajectoire, les bases terminales explorent différents types d'interaction et perturbent ainsi les pas voisins. Dans le chapitre 3 nous apporterons les preuves expérimentales que ces événements n'ont aucun fondement réaliste. En appliquant des contraintes sur les atomes lourds impliqués dans les liaisons hydrogènes des paires de bases terminales, toute interaction parasite disparaît et les problèmes de convergence sont résolus.

En résumé, CHARMM36 et Parmbsc0_{ezOLI} permettent désormais d'explorer bien mieux qu'avant les propriétés structurales et dynamiques de l'ADN. Néanmoins, d'autres efforts s'avèrent nécessaires pour améliorer la représentation de l'effet de séquence sur l'ADN et le comportement des paires de bases terminales.

2 Article 2

Simulations meet experiment to reveal new insights into DNA intrinsic mechanics

**Akli Ben Imeddourene, Ahmad Elbahnsi, Marc Guérout, Christophe Oguey,
Nicolas Foloppe, and Brigitte Hartmann**

***Plos Computational Biology* 2015, 11(12): e1004631.**

Simulations meet experiment to reveal new insights into DNA intrinsic mechanics

Akli Ben Imeddourene^{1,2}, Ahmad Elbahnsi^{1,3}, Marc Guérault⁴, Christophe Oguey³, Nicolas Foloppe^{5,*} and Brigitte Hartmann^{1,*}

¹ LBPA, UMR 8113, ENS de Cachan CNRS, 61 avenue du Président Wilson, 94235 Cachan cedex, France

² Université Pierre et Marie Curie, Paris, France

³ LPTM, Université de Cergy-Pontoise, 2 avenue Adolphe Chauvin 95031 Cergy-Pontoise, France

⁴ INSERM UMR S665, INTS, 6 rue Cabanel, 75015 Paris, France

⁵ 51 Natal Road, Cambridge CB1 3NY, UK

* Corresponding authors: nf_research@hotmail.co.uk (NF), bhartman@ens-cachan.fr (BH)

ABSTRACT

The accurate prediction of the structure and dynamics of DNA remains a major challenge in computational biology, due to the dearth of precise experimental information on DNA free in solution, and limitations in the DNA force-fields underpinning the simulations. A new generation of force-fields has been developed to better represent the sequence-dependent B-DNA intrinsic mechanics, in particular with respect to the BI \leftrightarrow BII backbone equilibrium, which is essential to understand the B-DNA properties. Here, the performance of MD simulations with the recent force-fields Parmbsc0_{OL1} and CHARMM36 was tested extensively against fresh NMR data collected on four DNA dodecamers involved in nucleosome positioning. We find impressive progress towards a coherent, realistic representation of B-DNA in solution, despite residual shortcomings. This improved representation allows new and deeper interpretation of the experimental observables, including regarding the behavior of facing phosphate groups in complementary dinucleotides, and their modulation by the sequence. It also provides the opportunity to extensively revisit and refine the coupling between backbone states and inter base pair parameters, which emerges as a common theme across all the complementary dinucleotides. In sum, the global agreement between simulations and experiment reveals new aspects of intrinsic DNA mechanics, a key component of DNA-protein recognition.

AUTHOR SUMMARY

The ability to simulate computationally the structure and dynamics of biomolecules is a major goal of structural biology. Such simulations require the calculation of the forces and energy of the system, typically with extensively parametrized functions called “force-fields”. Developing reliable force-fields has been very challenging for DNA, mainly because the simulations have to reproduce subtle, complex, sequence-dependent differences that also remain difficult to capture experimentally in solution. Here, we take advantage of an extensive set of fresh experimental (NMR) data gathered on selected DNA oligomers to test the performance of a new generation of force-fields for DNA simulations. Our results demonstrate impressive progress towards more realistic simulations of DNA. The agreement between experiment and simulations is now good enough to incite further interpretation of the experimental observables and yield new original insights into the intrinsic DNA mechanics. In sum, this work shows that reliable DNA simulations provide a much finer understanding of the structural and dynamical B-DNA behavior, which will be essential to account for DNA recognition by proteins.

INTRODUCTION

Binding of DNA to proteins or small molecules is modulated by subtle sequence-dependent variations inherent to the structure and dynamics of free DNA, which facilitate or disfavor the structural fit with cognate partners. Given the many DNA targets, a purely experimental characterization of their structure and dynamics is an enormous task. The structural biology of DNA would be greatly helped if one could describe and predict the sequence-dependent intrinsic mechanical and structural preferences of the double helix. That would pave the way to a fuller understanding of DNA malleability in direct and indirect readout.

Molecular Dynamics (MD) simulations in explicit solvent can potentially explore the properties of any B-DNA sequence of moderate length, considering the extensive sampling afforded by modern computational resources [1, 2]. However, MD simulations are only as reliable as the underlying energy model, typically treated with a classical force-field. Development of force-fields is complex, requires extensive efforts, and needs precise reference experimental data [3]. This latter requirement has been a complicating factor for DNA, given the paucity of reliable experimental data reflecting the fine structural details of DNA in solution [4-6]. The situation has improved in recent years, especially with respect to the DNA backbone, for which additional experimental information have been gathered from X-ray crystallography and NMR (see below). In response, force-field shortcomings regarding the DNA backbone were addressed [7-9], also motivated by the realization that the backbone is an essential ingredient of the intrinsic mechanical couplings in DNA.

Statistical analyses of X-ray structures of free B-DNA have unveiled that, among the five dihedral angles along the phosphate linkage, ϵ and ζ present bi-modal distributions [10-15], referred to as BI with ϵ/ζ :*trans/g-* and BII with ϵ/ζ :*g-/trans* [16, 17]. In contrast, α , β and γ appear to prefer overwhelmingly one conformation ($\alpha/\beta/\gamma$: *g-/t/g+*) [11, 13-15]. Importantly, the BI and BII conformers are associated to distinctive values of the helicoidal parameters, X-disp (base displacement), slide, roll and twist [11, 12]. In addition, the density of BI or BII phosphate groups in a window of 4 consecutive base pairs is coupled to the groove dimensions [18]. Hence, the modulation of B-DNA shape observed in X-ray structures is associated with the conformation of ϵ/ζ backbone dihedrals.

NMR solution studies echoed these findings and provided additional information about

the sequence dependent behavior of the BI and BII populations. In NMR, this equilibrium is reflected by the ^{31}P chemical shifts (δPs) [17, 19], which can be translated in terms of BII percentages [20, 21]. Correlations between NMR-measured δPs and internucleotide distances [20, 22] reflect the coupling between the backbone states and inter base pair parameters observed in X-ray structures. Consistency between NMR and X-ray results extends to the relation between the BII densities and the width of the minor groove, which can also be observed by NMR [23].

In addition, the compilation of a sizeable set of δPs documented the effect of B-DNA sequence on BII propensities [22], initially inferred from X-ray structures [11, 12]. Of the 10 unique complementary dinucleotides, CpG•CpG, CpA•TpG, GpC•GpC, GpG•CpC are characterized by BII percentages markedly higher than the average (21%); ApN•NpT (N: any base) and TpA•TpA can be globally considered as BI-rich; GpA•TpC is an intermediate case, with BII percentage only slightly lower than average. These intrinsic sequence-specific BII propensities in solution were summarized on a scale called TRX [22], by reference to the couplings with Twist, Roll and X-disp. TRX was recently validated further by a large set of δPs collected on new DNA oligomers [23].

This detailed information on DNA in solution offers a precious framework for testing and refining DNA force fields. Thus, the AMBER force fields Parm98 [24] and Parm99 [25] stabilized artefactual α/γ conformations that caused severe distortions in DNA [15, 26]. Such undesirable α/γ transitions were corrected in the subsequent potential Parmbsc0 [8]. First observed by NMR [27, 28] and then generalized and quantified in the TRX scale [22], the modulation of CpG BII propensity by the 3'- and 5'-neighbors was qualitatively retrieved by Parmbsc0 [29]. Also, the sensitivity of the BI \leftrightarrow BII equilibrium to the type of monovalent cation (K^+ , Na^+) demonstrated by NMR [30] was observed in Parmbsc0 simulations, which suggested a mechanism for this phenomenon [29]. Concerning the CHARMM family of force-fields, the early thorough systematic calibration of the DNA backbone torsional energetics for CHARMM27 [13, 31] prevented artefactual α/γ transitions and resulted in a force-field which treats B-DNA robustly [2, 32, 33]. Importantly, CHARMM27, like Parmbsc0, correctly represent the mechanical coupling between the backbone states and the helical parameters [5, 33, 34].

Nevertheless, the remaining shortcomings in Parmbsc0 MDs [1, 29, 35, 36] cannot be ignored, in particular regarding CpG, CpA and TpG that show a systematic deficit in BII with respect to the NMR data [20, 23, 27, 28, 37-42]. The CHARMM27 force-field did not neither reproduce the experimentally documented BII percentages [5, 34]. A simulation of the Drew-Dickerson dodecamer with Parmbsc0 [35] and a NMR/modeling study with CHARMM27 [43] also raised the issue of unrealistic BII propensities.

In response, two force-fields were recently conceived to improve the DNA backbone representation: Parmbsc0_{ε_{OLI}} [9] - derived from Parmbsc0 -, and CHARMM36 [7] - built on CHARMM27. Parmbsc0_{ε_{OLI}} and CHARMM36 were developed guided by DNA X-ray structures and a small set of BII percentages extracted from NMR. In initial tests with B-DNA, both force fields notably increased the sampling of the BII form compared to prior potentials [7, 9]. Since twist and groove shape are coupled to the BI \leftrightarrow BII equilibrium, the structural outcome obtained with Parmbsc0_{ε_{OLI}} significantly differs from that yielded by Parmbsc0 [9]. These initial tests are encouraging and call for a more systematic examination of the performance of these potentials, especially in the light of new experimental data not used to train the force-fields.

The present work exploits a wealth of fresh NMR data on the DNA backbone motions, to

thoroughly evaluate the performance of the Parmbsc0_{εζOLI} and CHARMM36 potentials. These data were collected on four DNA dodecamers [23], independent of those used to develop those force-fields. Together, the dodecamers cover a 39 bp segment in the 5' half of sequence 601, the best artificial sequence at forming nucleosome core particle [44], which is therefore important to understand how DNA is packaged. The TRX approach [22] combined with the analysis of the ³¹P chemical shifts of the four dodecamers [23] provides evidence that the intrinsic structural characteristics of the free sequence 601 largely account for its strong affinity for the histone core. In addition to their biological relevance, the 72 dinucleotides of the four dodecamers behave as expected from the TRX scale regarding the effect of sequence on the BII propensities [23]. These dodecamers and the attending experimental information are therefore ideally suited to evaluate Parmbsc0_{εζOLI} and CHARMM36, with emphasis on the representation of the BI ↔ BII equilibrium and the coupled helicoidal parameters. Importantly, we show how the improvements brought by these new potentials lead to new insights into DNA structure and dynamics, which are essentially consistent across the two force-fields.

The first step was to compare the BII percentages inferred from δP measurements to those generated by MDs. Although the simulated fine modulation by the sequence is not yet fully satisfactory, Parmbsc0_{εζOLI} and CHARMM36 represent the backbone behavior much more realistically than previous force fields. CHARMM36 in particular shows a very good ability to obtain BII-rich steps. This advance enabled to examine for the first time the conformational combinations corresponding to the states of facing phosphate groups, *i.e.* BI•BI, BI•BII, BII•BI and BII•BII. We find that the conformational states of the two facing phosphate groups of any complementary dinucleotide are not correlated in either Parmbsc0_{εζOLI} or CHARMM36 simulations. An important practical consequence is that the populations of the combinations of facing phosphates can be easily deduced from the overall individual BII populations inferred experimentally for every phosphate. This approach reveals that the four dodecamers contain a sizable number of steps where BII-containing states, BI•BII, BII•BI and BII•BII, dominate. Such quantification is critical for accurately describing the conformational landscape explored by the complementary dinucleotides, because the backbone combinations are tightly associated with helical parameters, as documented here. Overall, our results deepen our understanding of the intrinsic B-DNA mechanics, which is a key player in the indirect readout of DNA sequences by proteins.

RESULTS

Overview of the simulations.

Each of the four dodecamers (Table 1) was simulated with the Parmbsc0_{ε_{OLI}} [9] (P-MDs) or CHARMM36 [7] (C-MDs) force-field, resulting in a total of 8 MDs. The MDs of Oligo 1, 2 and 3 were 450ns each, while for Oligo 4 the trajectories were extended to one microsecond. Additional sampling was performed on Oligo 4 since its alternation of BI and BII-rich dinucleotides is especially relevant to test the convergence of backbone dynamics.

Table 1. Sequences of the four studied DNA dodecamers, constituents of the sequence 601.

Oligo 1	5'-TCGTAGCAAGCT-3'•5'-AGCTTGCTACGA-3'
Oligo 2	5'-GCTCTAGCACCG-3'•5'-CGGTGCTAGAGC-3'
Oligo 3	5'-CCGCTTAAACGC-3'•5'-GCGTTTAAGCGG-3'
Oligo 4	5'-CGCACGTACGCG-3'•5'-CGCGTACGTGCG-3'

During the present simulations, the dodecamers remained reassuringly close to canonical B-DNA. In both P-MDs and C-MDs the base pairs $N_2 \rightarrow N_{11} \bullet N_{14} \rightarrow N_{23}$ were stable, with ~99% of Watson-Crick pairing. The root mean square deviations (RMSDs) between canonical B-DNA and the simulated snapshots fluctuated around 2.6 ± 0.6 Å in P-MDs and 2.1 ± 0.5 Å in C-MDs (S1 Fig). The slightly larger RMSDs for P-MDs versus C-MDs gave the first indication of subtle differences between the force-fields. Then, we examined the five dihedral angles of the phosphodiester backbone, α , β , γ , ϵ and ζ . In both P- and C-MDs, $\alpha/\beta/\gamma$ conform to the canonical *g*-/*trans*/*g*+ pattern observed in free DNA [10, 11, 13-15]. The torsions ϵ and ζ undergo correlated motions, which define the BI and BII states (Fig 1). The convergence of the BII populations is of evident relevance, especially to compare simulated BII percentages to their experimental counterpart. Previous analyses of very long trajectories (up to ~45 μ s) with Parmbsc0 and CHARMM36 showed reasonable convergence of the fast motions (timescale < 100ns) on internal parts of DNA oligomers after only ~50ns [45]. A similar conclusion was drawn from μ s simulations with Parmbsc0, using as convergence criteria the average helical and backbone parameters [36]. These previous studies indicate that the timescale of the present MDs should be amply sufficient to investigate the backbone motions. Indeed, our results confirm this expectation, keeping in mind that constraints were applied to the terminal base-pairs to maintain their Watson-Crick base-pairing. A detailed justification of the protocol is given in Materials and Methods.

Fig 1. BI and BII conformations in the B-DNA backbone.

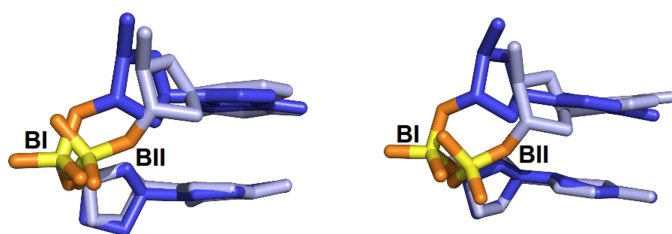


Illustration of the BI ($(\epsilon-\zeta) = -90^\circ$) and BII ($(\epsilon-\zeta) = +100^\circ$) phosphate linkage conformations with a GpC dinucleotide extracted from MDs carried out with the Parmbsc0 _{$\epsilon\zeta$ OLI} (left) or CHARMM36 (right) force fields.

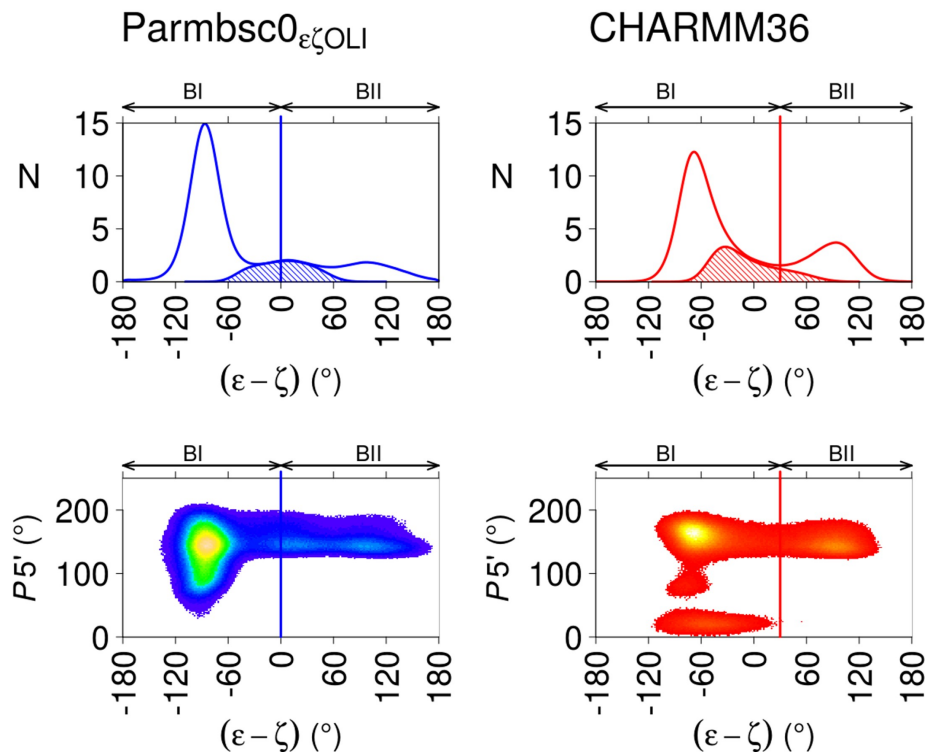
Clearly, the BII population of some phosphates did not converge over the first 50 ns (not shown), which may be considered as a reasonable equilibration time. Thus, for each phosphate, convergence was monitored by plotting its BII percentage over increasing trajectory lengths, from 150ns upwards (S2 Fig). For each Oligo treated with Parmbsc0 _{$\epsilon\zeta$ OLI} or CHARMM36, convergence of the BII populations was reached well within the simulation times, including for phosphates with high BII populations (S2 Fig). As a confirmation, the MDs of Oligo 4 extended to 1 μ s yielded the same BII percentages as the first part (50-450ns) of the simulation. Overall, we observe that the DNA backbone dynamics is essentially converged with MDs of several hundredth ns. This convergence timeframe is realistic considering that the phosphate groups undergo rapid (nano-picosecond timescale) conformational exchange according to NMR [46, 47]. Convergence in terms of sampling of the backbone states allows us to concentrate the following analysis on the influence of the force-fields.

Definition of the BI and BII states for MD analyses.

In free DNA X-ray structures the distribution of the pseudo-angle ($\epsilon-\zeta$) is characterized by two major peaks centered around ($\epsilon-\zeta$) = -90° (BI, ϵ in *trans* and ζ in *g*-) and ($\epsilon-\zeta$) = 90° , (BII, ϵ in *g*- and ζ in *trans*) (S3 Fig). Between these two maxima, a region covering ($\epsilon-\zeta$) values from -60 to $+70^\circ$ contains phosphate linkages with ϵ :*trans* typical of BI and ζ :*trans* typical of BII. Therefore this region may be considered ambiguous in terms of BI/BII categorization. The separation between BI and BII is commonly set at the minimum of the ($\epsilon-\zeta$) distribution, which is close to ($\epsilon-\zeta$) = 0 in the X-ray distribution (S3 Fig). BI and BII are thus usually characterized by negative and positive ($\epsilon-\zeta$) values, respectively.

The pattern observed in the X-ray structures for the ($\epsilon-\zeta$) distribution is globally preserved in P-MDs and C-MDs, while influenced by the force-fields (Fig 2). Thus, the operational definition of the BI and BII states in MDs must be carefully scrutinized, and possibly adapted. In addition to the ($\epsilon-\zeta$) histograms, the sugar populations in the *south*, *east* and *north* puckers (e.g. Oligo 4 sugars in S4 Fig) were considered, since this criterion is relevant to the definition of BI and BII. Indeed, crystallographic and NMR investigations established that BI is tolerant in terms of surrounding 5' and 3' sugar puckers, while BII is restricted to *south* puckers, especially with respect to 5' sugars [11, 14, 47].

Fig 2. Distribution of backbone ($\epsilon-\zeta$) values and 5' sugar behavior in MD snapshots.



Top panels: the frequencies (N) of ($\epsilon-\zeta$) values were extracted from P-MDs (left panel, blue line) and C-MDs (right panels, red line) snapshots. The BI and BII regions are indicated, in

accordance with the analysis presented in the main text. The hashed regions correspond to the distribution of ϵ/ζ :trans/trans, the categorization of which is ambiguous with regard to BI or BII. Bottom panels: pseudorotation phase angle (P) of the 5' sugar puckers as a function of $(\epsilon-\zeta)$ values ($^\circ$), extracted from P-MDs (left panel) and C-MDs (right panel). The sugar ring conformations were in north (pseudorotation phase angle 0 to 50°), east (50 to 120°) or south (120 to 220°). The $(\epsilon-\zeta)$ region assigned to BII does not contain 5' north sugars. The color gradient is indicative of the density, with the highest densities in yellow. The vertical lines indicate the $(\epsilon-\zeta)$ values used here to separate BI from BII in P- (blue lines) and C-MDs (red lines).

The $(\epsilon-\zeta)$ histogram of C-MDs has a minimum at $(\epsilon-\zeta) = 30^\circ$, located at a tail of the ϵ/ζ :trans/trans region (Fig 2). 5' south sugars are observable in both BI and BII regions; 5' east sugars fall in $(\epsilon-\zeta)$ from -110 to -50° , inside the conventional BI region; 5' north sugars are associated to a larger range of $(\epsilon-\zeta)$ values, but are suppressed above $(\epsilon-\zeta) = 30^\circ$ (Fig 2). This sugar behavior and the minimum of $(\epsilon-\zeta)$ at 30° offer an analogy with the X-ray observations, such that $(\epsilon-\zeta) = 30^\circ$ was deemed suitable to separate BI from BII with CHARMM36. Using instead the conventional cutoff ($\epsilon-\zeta = 0$) to separate BI from BII would not result in a dramatically different description, since the average BII population inferred with $(\epsilon-\zeta) > 0^\circ$ would only be 4% higher than that based on $(\epsilon-\zeta) > 30^\circ$. However, in view of the distribution of 5' north sugars, the criterion $(\epsilon-\zeta) > 30^\circ$ was preferred to define BII with CHARMM36. Incidentally, the present C-MD data confirm the strong association between BII and 5' south sugars.

With Parmbsc0 _{$\epsilon\zeta$ OLI}, no clear minimum is observed in the $(\epsilon-\zeta)$ distribution between the BI and BII peaks, which are separated by a flat $(\epsilon-\zeta)$ distribution (Fig 2). This intermediate region, centered around $(\epsilon-\zeta) = 0$, contains ϵ/ζ :trans/trans conformers (Fig 2), which represent 10% of the snapshots. In absence of a clear minimum in the $(\epsilon-\zeta)$ distribution, histograms of ϵ and ζ were considered. This approach was adopted in previous Parmbsc0 trajectories [48], where the BII linkages were defined relative to the minimum of the distribution. Here, with the minimum of the ζ distribution at $\zeta = 230^\circ$, this approach would designate as BII the range above $(\epsilon-\zeta) = -50^\circ$, a strongly negative value. Conversely, the transition from BI to BII would be at $(\epsilon-\zeta) = 40^\circ$ if chosen to be at the minimum of the ϵ histogram ($\epsilon = 240^\circ$). So, the ϵ and ζ histograms do not provide a coherent definition of BI and BII ranges for P-MDs here. In addition, the sugar dynamical regime is of little help since Parmbsc0 _{$\epsilon\zeta$ OLI} generates only a few north sugars (Fig 2 and S4 Fig). In absence of any convincing more specific rationale to assign the ϵ/ζ :trans/trans snapshots to either BI or BII with Parmbsc0 _{$\epsilon\zeta$ OLI}, we adopted the common BII definition, $(\epsilon-\zeta) > 0^\circ$. Such a decision is somewhat arbitrary but the uncertainty it introduces is limited. Indeed, shifting the $(\epsilon-\zeta)$ dividing value by 20° ($(\epsilon-\zeta) = -20^\circ$ or $+20^\circ$) only changed the BII population by $\pm 2\%$. In sum, BII percentages were extracted using $(\epsilon-\zeta) > 0^\circ$ for P-MDs and $(\epsilon-\zeta) > +30^\circ$ for C-MDs throughout this work.

These considerations illustrates the difficulty in defining BI and BII unambiguously, in a manner which would be meaningful and transferable across different structural models. It also

draws the attention to some differences between Parmbsc0_{εζOLI} and CHARMM36 regarding their representations of sugars and the (ε-ζ) distribution. Yet, the following sections show that a consistent overall picture emerges from CHARMM36 and Parmbsc0_{εζOLI}.

Overall BII populations from MD simulations compared to NMR.

The four dodecamers studied here correspond to 72 dinucleotide steps, excluding the terminal steps. The corresponding 72 ³¹P chemical shifts (δPs) were measured and converted to BII percentages, BII%_{from NMR}, using an empirical procedure based on a calibration involving a comparison of NMR and X-ray structural data [20] (see also Materials and Methods). In this procedure, δPs of pure BI and pure BII states are assumed to be sequence-independent, even if they could be modulated by the dinucleotide sequence, as suggested by a computational study [48]. However, previous studies showed that neglecting subtle sequence effect on δPs of pure BI and pure BII produced reasonable estimations [23, 49], for instance with points where BI and BII are expected to be equally populated [23]. Another indication of the protocol reliability is the consistency between the average BII percentage either derived from the average δPs of the 72 steps considered here (19% of BII, from δP_{av} = -4.20 ppm at 30°) or inferred from statistics of X-ray structures (20% of BII) [11].

A first test of the force fields is to compare the NMR-inferred and simulated BII populations, averaged on the 72 dinucleotides. The simulated overall average BII percentages are 11% in P-MDS and 18% in C-MDs. Thus, Parmbsc0_{εζOLI} somewhat underestimated the BII populations, as noted before [9]. The excellent agreement of the CHARMM36 value with experiment is an obvious improvement compared to CHARMM27, which severely underrepresented BII [5, 7].

That the force fields, in particular CHARMM36, produce overall BII population commensurate with experimental data is very encouraging, considering that the treatment of the backbone by previous force-fields fell outside the experimental range. Since the dinucleotides have markedly different propensities to populate BII [11, 22, 23, 48, 49], reproducing the sequence effects is a more stringent test of the force-fields, examined in the following.

Sequence-dependent BII propensities from simulations versus NMR.

A previous dataset of 323 measured δPs has established that the 16 dinucleotides composing B-DNA are associated with specific δP values [22]. Since δP translates into a BII propensity it implies that the BI/BII populations are primarily controlled by the dinucleotide sequence [22]. The additional 72 δPs considered here conform to this sequence pattern, validating the notion of dinucleotide-specific BII propensity [23]. Thus, the sequence-dependent BII populations derived from δPs provide a rare opportunity to test the sequence-dependent behavior of DNA force-fields in solution. One notes that adjustments made to Parmbsc0_{εζOLI} and CHARMM36 to increase the BII populations were not tailored depending on the base sequence, in contrast with, for instance, the CMAP approach [50]. In other words, the same backbone force-field parameters are applied to any sequence. Therefore, differences in the backbone behavior during simulations can only be ascribed to intrinsic sequence-dependent properties.

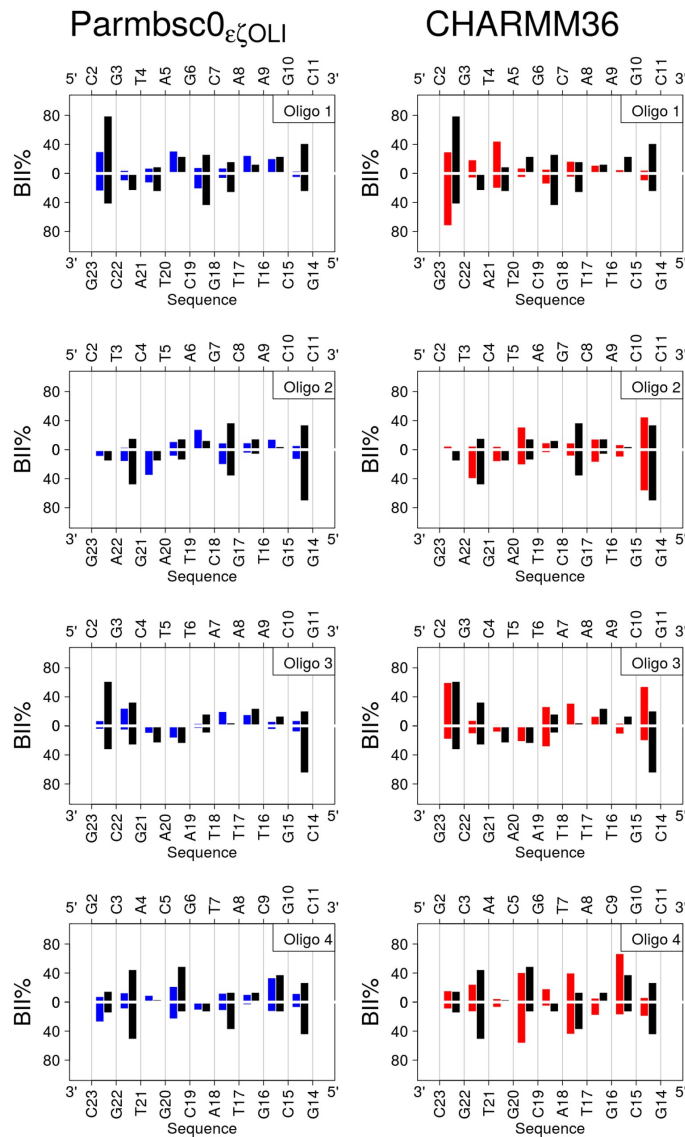
To examine whether Parmbsc0_{εζOLI} and CHARMM36 reproduce the effect of sequence on BII populations, the simulated BII percentages were compared to their experimental counterparts, considering the individual phosphates (BII%_{from MD} versus BII%_{from NMR}, given in S1 Table). BII%_{from MD} and BII%_{from NMR} are overall moderately correlated (Table 2 and S5 Fig). The simulated BII percentages of half of the 72 steps (53% for both P-MDs and C-MDs) are within BII%_{from NMR} ±10%, where the 10% interval corresponds to the tolerance allowed around the NMR-based BII percentages (see Materials and Methods). The comparison between BII%_{from NMR} and BII%_{from MD} is shown in Fig 3 for each non-terminal phosphate of the four dodecamers.

Table 2. BII percentages from NMR compared to their simulated counterparts.

	BII%			
	P-MDs		C-MDs	
	CC	Δ_{av}	CC	Δ_{av}
all	0.42	14 (14)	0.47	14 (12)
Oligo 1	0.62	14 (13)	0.33	18 (14)
Oligo 2	0.31	14 (15)	0.75	10 (8)
Oligo 3	0.32	14 (17)	0.49	14 (12)
Oligo 4	0.36	13 (14)	0.35	16 (13)

The linear correlation coefficient (CC) were calculated between BII percentages (BII%) inferred from experimental δP s and those extracted from the simulations with Parmbsc0_{εζOLI} (P-MDs) and CHARMM36 (C-MDs). Δ_{av} represents the average difference between the simulated and experimental BII% values; standard deviations are in brackets.

Fig 3. Comparison between simulated and experimental BII percentages along the four dodecamers.



BII percentages (BII%) were plotted along the two complementary strands of each dodecamer sequence. BII% were extracted from P-MDs (left panels, blue) and C-MDs (right panels, red), or inferred from the 72 δ Ps collected by NMR (black). The error on BII% from NMR was estimated to be $\pm 10\%$.

A more detailed analysis, in particular on the dinucleotides present in several occurrences

in the dodecamers, shows that Parmbsc0_{εζOLI} and CHARMM36 correctly reproduce the low or moderate BII%_{from NMR} (< 20%) of CpT, GpT, ApC, ApG, ApA and TpT (Fig 3 and Table 3 for the steps present in several occurrences in the dodecamers). However, this overall agreement suffers some shortcomings. The BII population of TpA tends to be either underestimated or overestimated with Parmbsc0_{εζOLI} and CHARMM36, respectively. GpC steps in BII are quasi systematically underestimated by both force fields, as well as one of the three CpA•TpG in Oligo 4. Parmbsc0_{εζOLI} generated too low BII percentages on CpC and GpG in Oligo 2 and most CpG (seven on a total of ten). In C-MDs, the representation of CpC, GpG and CpG is reasonable whereas inversions of BII% occurs at two complementary CpG•CpG steps. Indeed, the NMR gives asymmetrical BII percentages for C₂pG₃•C₂₂pG₂₃ in Oligo 1, with 79% of BII for C₂pG₃ and 42% for C₂₂pG₂₃. C-MD gave the reverse, with 30% and 75% of BII for C₂pG₃ and C₂₂pG₂₃, respectively. A similar situation arises for C₁₀pG₁₁•C₁₄pG₁₅ in Oligo 3.

Table 3. BII percentages from NMR compared to their simulated counterparts for selected steps.

	N	BII% _{from NMR}	BII% _{from P-MDs}	BII% _{from C-MDs}
CpT	6	0 (0)	0 (0)	4 (2)
GpT	6	0 (0)	3 (1.5)	14 (5.5)
TpT	4	0 (0)	0 (0)	2 (1)
ApC	6	12 (7)	10 (3)	5 (1)
ApA	4	16 (10)	19 (4)	18 (9)
TpA	8	17 (9.5)	9 (4)	32 (10)
ApG	6	19 (5)	22 (11)	8 (5)
GpC	12	31 (10)	12.5 (8)	10 (4)
CpG	10	41 (22)	17 (10.5)	43 (21)

The BII percentages inferred from experimental δP s or extracted from the simulations with Parmbsc0_{εζOLI} (P-MDs) and CHARMM36 (C-MDs) are reported for the steps present in more than 3 occurrences (N) in the dodecamers. The values are averaged, with standard deviations in brackets.

Our results confirm that adjustments specifically aimed at enhancing access to the BII state produce convincing, positive effects, especially perceptible in CHARMM36. That the increase in simulated BII populations is not distributed uniformly along the sequences (Fig 3) is not trivial since, as noted above, the computational models were not parametrized to reproduce the BII% for specific dinucleotides, but were only adjusted to be generically more permissive to BII. Admittedly, discrepancies still exist between the experimental sequence effect on the BI↔BII equilibrium and Parmbsc0_{εζOLI} or CHARMM36. However, an essential point is that the simulations are now sufficiently BII-rich to extend the analysis to aspects of the backbone dynamics that eludes experimental approaches.

New insights from the simulations: independence of the states of facing phosphate groups.

The phosphate groups facing each other across the strands can adopt homogeneous combinations, BI•BI or BII•BII, or hybrid combinations, BI•BII or BII•BI (denoted here BI•BII|BII•BI, where the vertical bar means logical “or”). The populations of these combinations are especially meaningful from the point of view of B-DNA mechanics, because inter base pair parameter values are associated to the conformational states of two facing phosphate linkages [11, 12, 20, 35, 51-54], as also addressed below.

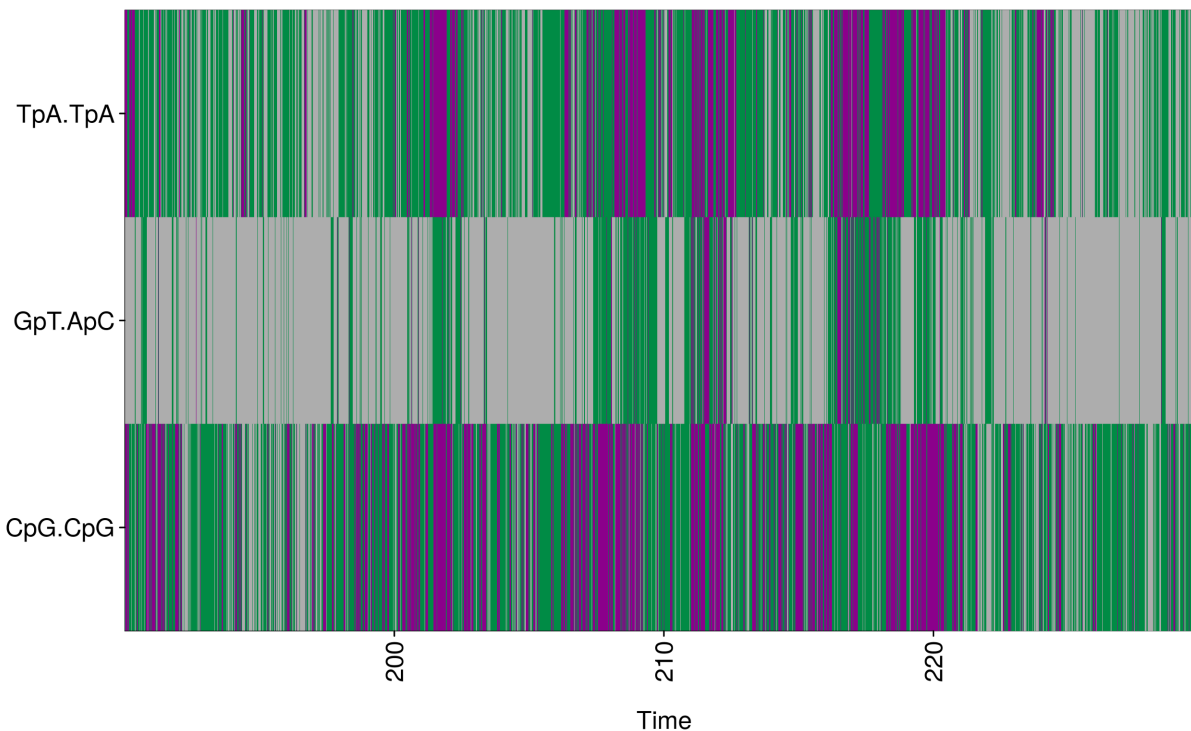
The behavior of facing phosphate linkages cannot be deduced from δP measurements, which report time and ensemble-averaged BII percentages for individual phosphates. The present simulations offer the opportunity to inspect the dynamic behavior of phosphate linkages in complementary dinucleotides and to estimate possible correlation. Indeed, several steps in C-MDs, in particular CpG•CpG, CpC•GpG and TpA•TpA, adopt the three combinations, BI•BI, BI•BII|BII•BI or BII•BII (Table 4 and Fig 4). In P-MDs, BI•BI and BI•BII|BII•BI are also frequently observed, but the BII•BII populations are almost inexistent (Table 4), consistent with Parmbsc0_{OL1} generating fewer BII conformers than CHARMM36.

Table 4. Conformational combinations of facing phosphates in MDs.

	N	(BI•BI)%		(BI•BII BII•BI)%		(BII•BII)%	
		P-MDs	C-MDs	P-MDs	C-MDs	P-MDs	C-MDs
CpG•CpG	5	67 (18)	27 (7)	31 (17)	58 (5)	1 (1)	14 (6)
CpC•GpG	1	82	26	18	47	1	27
TpA•TpA	4	83 (7)	46 (11)	17 (7)	44 (7)	0	10 (4)
GpC•GpC	6	76 (9)	81 (4)	23 (9)	18 (4)	1 (1)	1 (0)

The percentages of the conformational combinations of facing phosphate linkages, BI•BI, BI•BII|BII•BI, and BII•BII, are reported for selected steps with substantial BII% during either P-MDs or C-MDs (see Fig 3). The values corresponding to CpG•CpG, TpA•TpA, and GpC•GpC steps, present in several occurrences (N) in the dodecamers, are averaged, with standard deviations in brackets.

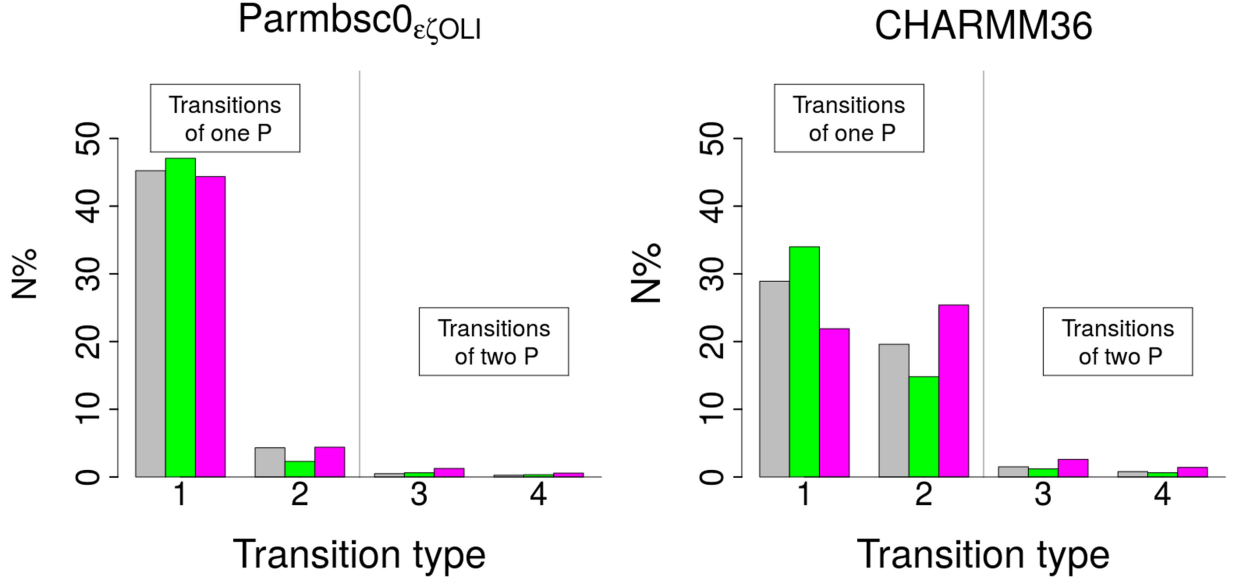
Fig 4. Conformational combinations of facing phosphate linkages illustrated with CHARMM36.



In C-MDs, the BII-rich facing phosphate linkages adopt three conformational combinations, BI•BI (grey), BI•BII|BII•BI (green) and BII•BII (violet). The course of these combinations versus time (ns) is illustrated for facing phosphate groups of Oligo 4, two BII-rich steps, $C_{5p}G_6 \bullet C_{19p}G_{20}$ and $T_{7p}A_8 \bullet T_{17p}A_{18}$, and one BI-rich step, $G_{6p}T_7 \bullet A_{18p}C_{19}$. For clarity, only a small part of the trajectory is shown here.

Fig 5 illustrates the statistics of the transitions between the facing phosphate combinations for steps that adopt BII•BII in both P-MDs and C-MDs. The same result holds for any other complementary step investigated here in which the facing phosphates undergo BI \leftrightarrow BII transitions. With both force fields, the large majority of the transitions between BI•BI, BI•BII|BII•BI and BII•BII involves only one of the two facing phosphates (BI•BI \leftrightarrow BI•BII or BII•BI; BI•BII or BII•BI \leftrightarrow BII•BII). BI•BII|BII•BI \leftrightarrow BII•BII are infrequent in P-MDs, the BII•BII state being poorly populated. In both P-MDs and C-MDs, the simultaneous transitions of two phosphate states (BI•BII \leftrightarrow BII•BI; BI•BI \leftrightarrow BII•BII) are very rare, representing at most 5% of the total number of transitions (Fig 5).

Fig 5. Statistics on the transitions between the conformational states of facing phosphate groups.



The transitions between facing phosphate group combinations were analyzed for $C_{5p}G_6 \cdot C_{19p}G_{20}$ (grey) and $T_{7p}A_8 \cdot T_{17p}A_{18}$ (green) in Oligo 4, and $C_{10p}C_{11} \cdot G_{14p}G_{15}$ (pink) in Oligo 2, from P-MDs (left) and C-MDs (right). These steps were chosen because they adopt BII•BII in the simulations. N% is the percentage of a transition type relative to the total number of transitions. The transition types are labelled as follow: 1: $BI \cdot BI \rightarrow BI \cdot BII | BII \cdot BI$ or the inverse, $BI \cdot BII | BII \cdot BI \rightarrow BI \cdot BI$; 2: $BI \cdot BII | BII \cdot BI \rightarrow BII \cdot BII$ or $BII \cdot BII \rightarrow BI \cdot BII | BII \cdot BI$; 3: $BI \cdot BII \rightarrow BII \cdot BI$ or $BII \cdot BI \rightarrow BI \cdot BII$; 4: $BI \cdot BI \rightarrow BII \cdot BII$ or $BII \cdot BII \rightarrow BI \cdot BI$.

The populations of $BI \cdot BI$, $BI \cdot BII | BII \cdot BI$ and $BII \cdot BII$ can be addressed with simple elements from probability theory, summarized here before comparison to the MD data. $P_i(BII)$ is the probability that phosphate i is in BII, the complementary event has probability $P_i(BI) = 1 - P_i(BII)$. The states of facing phosphate pairs i, j are characterized by pair probability distributions, $P_{ij}(BI \cdot BI)$, $P_{ij}(BI \cdot BII)$, $P_{ij}(BII \cdot BI)$ and $P_{ij}(BII \cdot BII)$. Because the facing phosphate groups are either BI or BII, the probabilities satisfy the relations:

$$P_{ij}(BI \cdot BI) + P_{ij}(BI \cdot BII) = P_i(BI)$$

$$P_{ij}(BI \cdot BI) + P_{ij}(BII \cdot BI) = P_j(BI)$$

$$P_{ij}(BII \cdot BI) + P_{ij}(BII \cdot BII) = P_i(BII)$$

$$P_{ij}(BI \cdot BII) + P_{ij}(BII \cdot BII) = P_j(BII).$$

Summing the last two equations gives:

$$[P_{ij}(\text{BII} \bullet \text{BI}) + P_{ij}(\text{BI} \bullet \text{BII})] + 2 P_{ij}(\text{BII} \bullet \text{BII}) = [P_i(\text{BII}) + P_j(\text{BII})] \quad (1)$$

The first term on the left in (1) is the probability of $\text{BI} \bullet \text{BII} | \text{BII} \bullet \text{BI}$, $P_{ij}(\text{BII} \bullet \text{BI} | \text{BI} \bullet \text{BII})$. Note that here $P_{ij}(\text{BII} \bullet \text{BI} | \text{BI} \bullet \text{BII})$ does not denote a conditional probability, but simply the probability of states $\text{BII} \bullet \text{BI}$ or $\text{BI} \bullet \text{BII}$. So equation (1) is equivalent to

$$P_{ij}(\text{BII} \bullet \text{BI} | \text{BI} \bullet \text{BII}) + 2 P_{ij}(\text{BII} \bullet \text{BII}) = P_i(\text{BII}) + P_j(\text{BII}). \quad (1')$$

Equation (1') is general, as it follows directly from the definitions and it does not rely on any assumption about the independence (correlation) of facing phosphate groups.

One now examines the case when the states of the two facing phosphate groups are independent of each other. Then, the pair probabilities factorize: $P_{ij}(b \bullet b') = P_i(b) P_j(b')$, where b and b' stand for any of the phosphate states. In particular, we have

$$P_{ij}(\text{BII} \bullet \text{BII}) = P_i(\text{BII}) P_j(\text{BII}) \quad (2)$$

$$P_{ij}(\text{BII} \bullet \text{BI} | \text{BI} \bullet \text{BII}) = P_i(\text{BII}) + P_j(\text{BII}) - 2P_i(\text{BII}) P_j(\text{BII}). \quad (3)$$

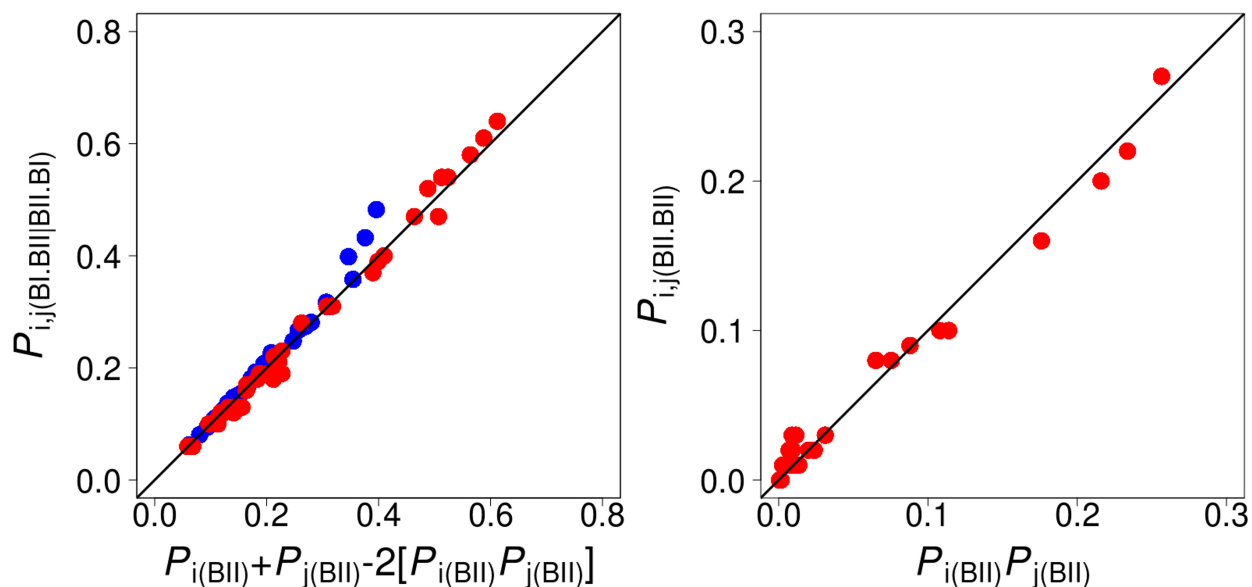
Equation (3) follows from (1') and (2) and from the relations: $P_i(\text{BI}) = 1 - P_i(\text{BII})$ and $P_j(\text{BI}) = 1 - P_j(\text{BII})$.

Equations (2) and (3) mean that, under the assumption of statistical independence of the two individual facing phosphates, the knowledge of the single phosphate probabilities $P_i(\text{BII})$ and $P_j(\text{BII})$ is sufficient to find the probabilities of $\text{BII} \bullet \text{BI} | \text{BI} \bullet \text{BII}$, $\text{BII} \bullet \text{BII}$, and then also of $\text{BI} \bullet \text{BI}$ by using equation:

$$P_{ij}(\text{BI} \bullet \text{BI}) = 1 - [P_{ij}(\text{BII} \bullet \text{BI} | \text{BI} \bullet \text{BII}) + P_{ij}(\text{BII} \bullet \text{BII})].$$

The next step was to test the possibility of uncorrelated facing phosphates against data collected from the MDs. Thus, $P_{ij}(\text{BII} \bullet \text{BI} | \text{BI} \bullet \text{BII})$, $P_{ij}(\text{BII} \bullet \text{BII})$, $P_i(\text{BII})$ and $P_j(\text{BII})$ were evaluated as the proportions of these states in the MD trajectories; $P_{ij}(\text{BII} \bullet \text{BI} | \text{BI} \bullet \text{BII})$ was compared to $[P_i(\text{BII}) + P_j(\text{BII}) - 2P_i(\text{BII}) P_j(\text{BII})]$ in P-MDs and C-MDs; $P_{ij}(\text{BII} \bullet \text{BII})$ was compared to $[P_i(\text{BII}) P_j(\text{BII})]$ in C-MDs only, since they generate sizable $\text{BII} \bullet \text{BII}$ populations in contrast with P-MDs. The agreement between the compared quantities is clearly visible in Fig 6, with correlation coefficients of 0.99. That is, the distribution of BII steps between the $\text{BI} \bullet \text{BII} | \text{BII} \bullet \text{BI}$ and $\text{BII} \bullet \text{BII}$ combinations in complementary dinucleotide matches equations (2) and (3) very well. Thus, the conformational states of the facing phosphates are statistically independent of each other.

Fig 6. Facing phosphate states pair correlations.



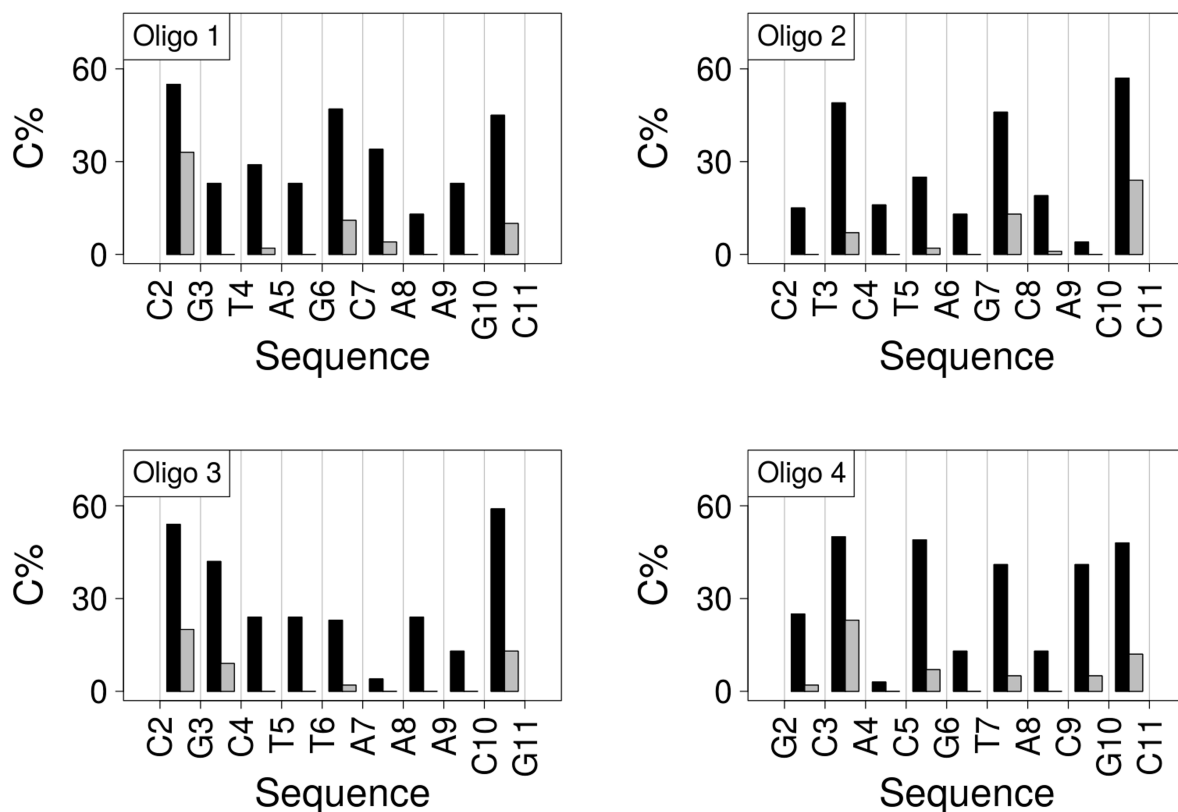
The facing phosphate states $BI\bullet BII|BII\bullet BI$ and $BII\bullet BII$ have probabilities $P_{ij}(BII\bullet BI|BI\bullet BII)$ (left panel) and $P_{ij}(BII\bullet BII)$, (right panel), respectively, extracted from P-MDs (blue) and C-MDs (red) for each complementary step in the dodecamers. $P_{ij}(BII\bullet BII)$ versus $P_{i(BII)}P_{j(BII)}$ is not reported for P-MDs because of the much reduced $BII\bullet BII$ populations in these simulations. These P_{ij} probabilities are compared to expressions containing the individual BII probabilities, $P_{i(BII)}$ and $P_{j(BII)}$, under the assumption that the conformational states of facing phosphate groups are independent of each other (equation (3) for $P_{ij}(BII\bullet BI|BI\bullet BII)$ and equation (2) for $P_{ij}(BII\bullet BII)$, see main text). The diagonal lines correspond to $y = x$.

In sum, the ability of both Parmbsc0_{εζOLI} and CHARMM36 to generate phosphates visiting BII enabled to gain new insights into their dynamics and populations. Thus, simultaneous transitions of two facing phosphate groups are very rare. The two force-fields unambiguously support the notion of statistical independence of the conformational states of individual, facing phosphates. This means that the populations of the three combinations of facing phosphates can be simply expressed from the BII propensities of individual phosphate groups, in particular from experimental data, as developed in the next section.

Quantifying the facing phosphate combinations from NMR.

Considering that the notion of uncorrelated facing phosphate is convincing, the probabilities of states BI•BII|BII•BI and BII•BII were calculated using equations (2) and (3), respectively, and equating $P_i(\text{BII})$ and $P_j(\text{BII})$ to $P_{i \text{ from NMR}}(\text{BII})$ and $P_{j \text{ from NMR}}(\text{BII})$ (equivalent to $\text{BII}\%_{\text{from NMR}}$ given in S1 Table). This has the advantage to use the experimental data directly (δP -derived BII percentages) to quantify the phosphate states, bypassing the limitations in the simulated estimates the phosphate state populations. The resulting experimentally inferred BI•BII|BII•BI and BII•BII populations along the four dodecamers are shown in Fig 7 and the values are given in S2 Table. According to this approach, most CpG•CpG and GpC•GpC, as well as the only CpC•GpG, are characterized by high percentages (45% and more) of BI•BII|BII•BI (Fig 7). CpG•CpG in Oligos 1 and 3, CpC•GpG in Oligo2 and CpA•TpG in Oligo 4 are in addition more than 20% in BII•BII (Fig 7). Overall, BI•BI is not the most frequent state in 12 steps, out of a total of 36, in the four dodecamers.

Fig 7. Quantification of the conformational combinations of independent facing phosphates based on NMR data.



The percentages of BI•BII|BII•BI (black bars) and BII•BII (grey bars) combinations (C%) of facing phosphates are plotted along the four dodecamer sequences. These percentages were calculated in the regime of independence of the conformational states of facing phosphates with equations (2) and (3) and the NMR data. The values are given in S1 Table.

As seen above (Fig 3), the individual BII percentages extracted from MDs differ from those inferred from NMR; accordingly, the corresponding respective populations of BI•BII|BII•BI and BII•BII are not identical. However, the match between C-MD and experimentally inferred data is reasonable (S7 Fig), with correlation coefficients of 0.62 for BI•BII|BII•BI and 0.57 for BII•BII. So, CHARMM36 appears to represent the sequence-dependent behavior of the pairs of facing phosphates better than that of individual phosphates (Table 2). This improvement reflects in part compensatory effects between the two strands of CpG•CpG steps, in which the asymmetric individual BII percentages are inversed in C-MDs and NMR (see the above section “Sequence-dependent BII propensities from simulations versus NMR”).

Overall, the realization that the states of facing phosphates are independent enables to derive their populations from δP -based BII percentages. Applying this approach reveals that all the complementary dinucleotides in the four dodecamers populate both BI•BI and BI•BII|BII•BI, some of them also display significant percentages of BII•BII (Fig 7 and S2 Table). This prevalence of BII-containing steps is of real importance for the DNA intrinsic mechanics, as examined next.

Conformational combinations of facing phosphate and inter base pair parameters.

BI•BI, BI•BII|BII•BI or BII•BII are associated to different values of slide, roll and twist in X-ray structures [11, 12, 34]. However, the requirement to select only very high resolution X-ray structures to ensure the accuracy of backbone dihedral angles [55] drastically limits the data for analysis. A previous study [22] underlined that BII conformers in such X-ray structures occur almost exclusively in CpG, CpA, TpG, GpG, and GpC; furthermore, the BII•BII combination was only observed in CpA•TpG. The improved representation of the DNA backbone with Parmbsc0_{εζOLI} and CHARMM36 offers the opportunity to broaden the analysis of the helicoidal parameters associated to the facing phosphate combinations for a larger variety of complementary dinucleotides than in X-ray datasets. Consistent results between both force fields would of course strengthen the conclusions.

The mean values of the six inter base pair parameters (shift, slide, rise, slide, roll and twist) were calculated for each conformational combination of the facing phosphate groups, after merging all equivalent conformational combinations across complementary steps. Slide, roll and twist are found very sensitive to the facing backbone conformational combinations (Table 5) contrary to invariant shift and tilt (S3 Table). As in a previous study using Parmbsc0 [35], rise variations are observed, but the change between BI•BI and BII•BII does not exceed 0.2 Å in both P-MDs and C-MDs (S3 Table).

Table 5. Inter base pair parameters variations associated to the conformational states of the facing phosphate groups.

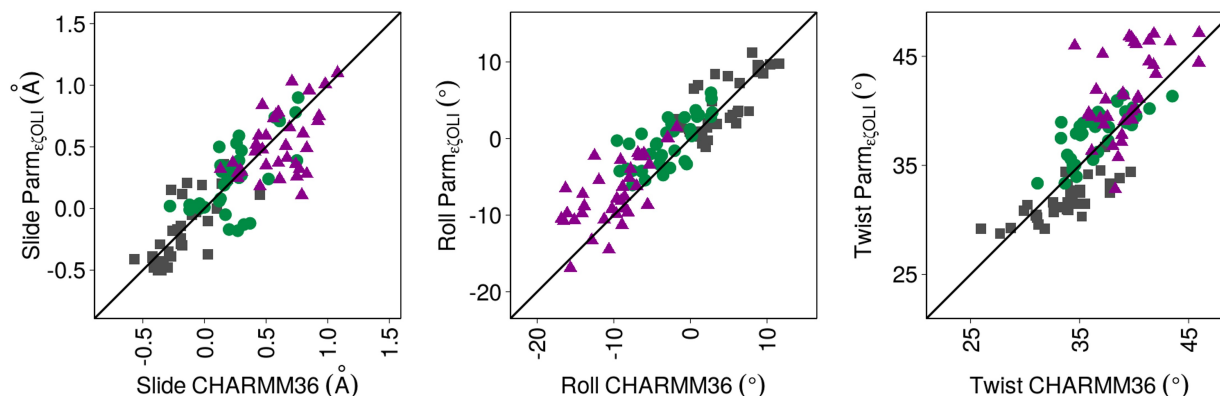
	Slide (Å)		Roll (°)		Twist (°)	
	P-MDs	C-MDs	P-MDs	C-MDs	P-MDs	C-MDs
BI•BI	-0.2 (0.3)	-0.2 (0.2)	4 (3.7)	3.8 (3.7)	31.9 (1.9)	33.7 (3.0)
BI•BII BII•BI	0.2 (0.3)	0.2 (0.2)	-0.5 (3.3)	-3.0 (3.6)	37.9 (2.3)	36.5 (2.6)
BII•BII	0.6 (0.3)	0.6 (0.2)	-7 (4.3)	-9.9 (4.1)	41.7 (3.8)	39.3 (2.6)
$\Delta(\text{BII} \cdot \text{BII} - \text{BI} \cdot \text{BI})$	+0.8	+0.8	-11	-13.7	+9.8	+5.6

Values of Slide, Roll and Twist generated by Parmbsc0_{εζOLI} (P-MDs) and CHARMM36 (C-MDs) for the three conformational combinations of facing phosphate groups, extracted after merging data from all simulated complementary dinucleotides in the four dodecamers. Standard deviations are in brackets. The last row reports the difference between the inter base pair parameters in the two extreme situations, BI•BI and BII•BII.

It is striking that there is almost quantitative agreement on the changes of slide, roll and twist across BI•BI, BI•BII|BII•BI and BII•BII with both force-fields (Table 5). Yet, CHARMM36

does not increase the twist as much as $\text{Parmbsc0}_{\text{e}\zeta\text{OLI}}$ from BI•BI to BII•BII. Considering the 36 individual complementary dinucleotides of the four oligomers (Fig 8) confirms this concordance. Only isolated departures between $\text{Parmbsc0}_{\text{e}\zeta\text{OLI}}$ and CHARMM36 appear when the variation of the helical parameters is examined in individual complementary steps (examples in S6 Fig). In P-MDs, the rolls of TpA•TpA are systematically $5\pm 2.5^\circ$ larger than in C-MDs, for all backbone combinations; the twist of CpG•CpG and CpA•TpG in BII•BII is $6.5\pm 1.5^\circ$ higher in P-MDs than in C-MDs.

Fig 8. Slide, Roll and Twist values generated by $\text{Parmbsc0}_{\text{e}\zeta\text{OLI}}$ and CHARMM36 for individual complementary dinucleotide steps, versus the facing phosphate combinations.

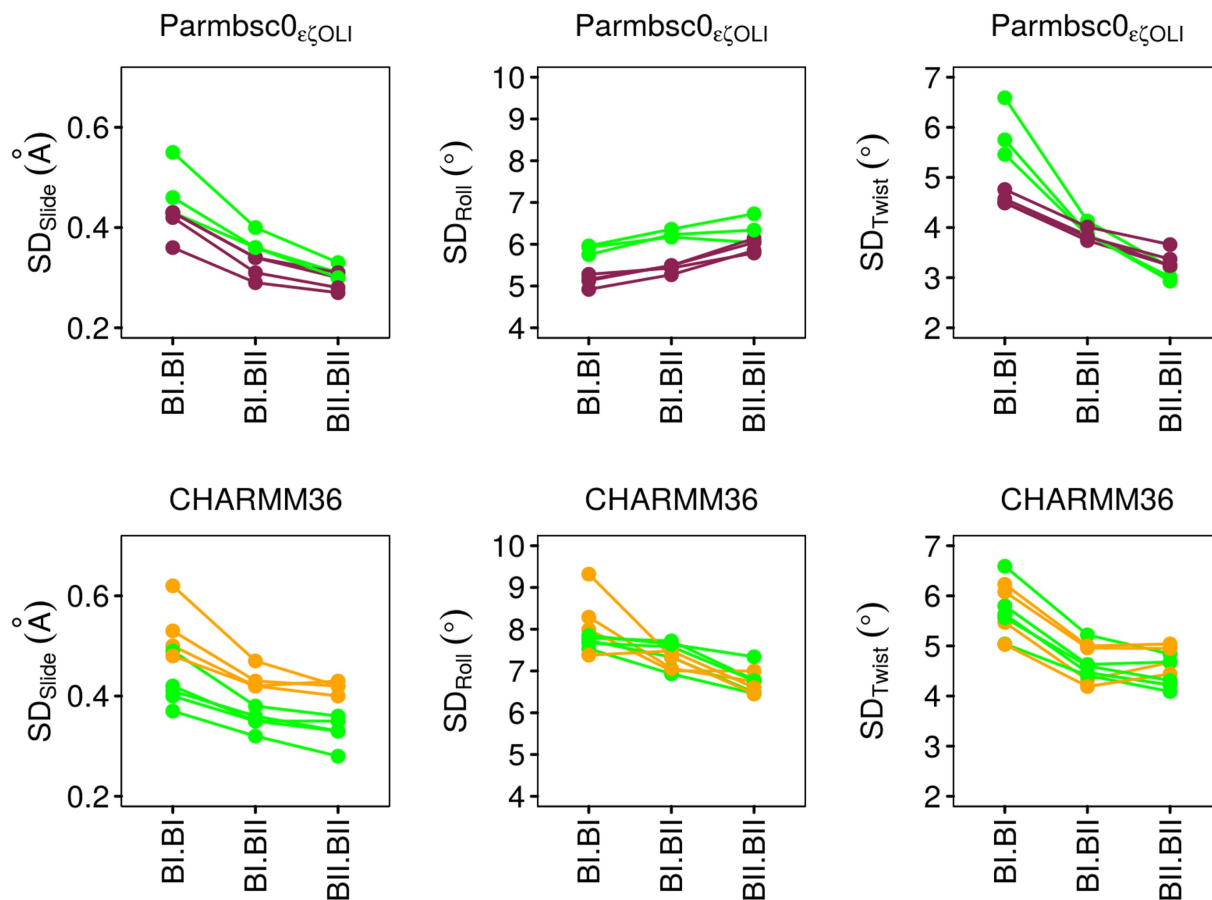


The mean values of Slide, Roll and Twist were calculated over the MDs for each of the 36 complementary dinucleotides of the four dodecamers, categorized according to the BI•BI (grey squares), BI•BII/BII•BI (green circles) and BII•BII (violet triangles) combinations of their facing phosphate groups. The data were extracted from P-MDs and C-MDs, and time-averaged for each conformational combination. The standard deviations are 0.6 for slide, $\sim 7^\circ$ for roll and 6° for twist, with both force-fields. The correlation coefficients are 0.84 (P-slide versus C-slide), 0.90 (P-roll versus C-roll) and 0.78 (P-twist versus C-twist). The diagonal lines correspond to $y = x$.

The MD results not only systematically documented the couplings, but they enabled comparison of the variability (standard deviations) of the helicoidal parameters for BI•BI, BI•BII/BII•BI and BII•BII. In both P-MDs and C-MDs, the slide and twist variabilities are greater in BI•BI compared to BI•BII/BII•BI or BII•BII (Fig 9). A similar, but attenuated, trend is observed for the roll in C-MDs (where the roll standard deviations are higher than in P-MDs). Thus, the simulations suggest that BII containing combinations are stiffer than BI•BI, at least for slide and twist. This, combined with the suppression of north sugars in 5' of the BII linkages, might entropically disfavor the BII conformers. However, other contributions will affect the net

balance of the BI \leftrightarrow BII equilibrium. Indeed, the above quantitative analysis makes clear that BII is frequently populated, and is the dominant conformer at some base steps.

Fig 9. Variability of helical parameters depending on conformational combinations of facing phosphate linkages.



The standard deviations of slide (SD_{Slide}), roll (SD_{Roll}) and twist (SD_{Twist}) associated to the three possible combinations of facing phosphates are shown for representative steps, CpG•CpG in P-MDs and C-MDs of Oligos 1, 3 and 4 (green), GpC•GpC in P-MDs of Oligos 1, 2, 3 and 4 (violet) and TpA•TpA in C-MDs of Oligos 1, 3 and 4 (orange). Steps with the largest sampling of BII•BII (1.1 to 2.9%), were selected from P-MDs and C-MDs (7.5 to 21.8%). Top panels: P-MDs; bottom panels: C-MDs.

The concordant results from P-MDs and C-MDs considerably strengthen and extend the view of the couplings between the facing phosphate states and inter base pair parameters gleaned from X-ray structures. Here, MDs inform about the behavior of a large range of dinucleotides, comprising those that are moderately or even barely propitious to BII. They reveal a general mechanical property of free DNA. Thus, BII containing complementary dinucleotides are characterized by more positive slide, more negative roll and higher twist than those in BI. As most steps have access to the BI•BII|BII•BI states (see the preceding section), the DNA deformation cost upon protein binding could be less important than expected when BII-like features are required for recognition. This can be illustrated by the TTAAA sequence in Oligo 3. This segment is considered as one of the strongest anchoring points in the nucleosome assembly [56-60], by forming multiple interactions with histones H3 and H4. In the X-ray structures of nucleosome containing the sequence 601 (PDB entries 3ZL0, 3ZL1[59] and 3MVD [57]), TTAAA•TTTAA displays rather variable but globally negative rolls ($-7 \pm 6^\circ$). According to both NMR and MDs, in their free state, these steps are mainly in BI•BI, associated to rolls of $4.4 \pm 3^\circ$. However, they also explore the BII•BI|BI•BII states, with rolls of $-2.5 \pm 1^\circ$. So, the free TTAAA sequence spontaneously visits conformations closer than expected to its bound counterpart.

DISCUSSION

Assessing the extent to which MD simulations correctly represent B-DNA structural features in solution, their sequence dependency and populations remains an ongoing challenge and a necessary step to gain confidence in the role that DNA simulations may play in biophysics and structural biology. Part of the difficulty is to obtain experimental data in solution, suitable for comparison with simulations. Here, Parmbsc0_{ε_{OL}} [9] and CHARMM36 [7], specifically developed to improve the representation of the DNA backbone, were tested with respect to the sequence-specific BI and BII populations in four dodecamers, derived from ³¹P chemical shifts (δPs) [20].

The results show that the Parmbsc0_{ε_{OL}} and CHARMM36 potentials produce substantial BII populations, closer to their experimentally inferred counterpart than those obtained with preceding force fields [1, 5, 35, 36]. Many simulated BII propensities of the four dodecamers compare satisfactorily to experiment, a very encouraging achievement. This provides the foundation to understand the factors underpinning the differentiated BI ↔ BII equilibrium behavior across base steps. In particular, the context may be better adapted to investigate the quantum-mechanical origin of the phosphate chemical shifts [48].

However, the experimental sequence effect on BII propensities is still imperfectly reproduced by simulations, each force field displaying its own weaknesses. Parmbsc0_{ε_{OL}}, as reported by its developers [9], globally underestimates the BII propensities. The CHARMM36 biases include generating too high BII percentages on TpA or, conversely, suppressing the BII character of GpC and some CpA and TpG. The procedure translating experimental δP to BII% is not devoid of uncertainties [48], but they would not account for the most severe discrepancies. For instance, the simulated TpA BII-richer than GpC is clearly inconsistent with both NMR and X-ray data [11, 20, 22]. There was no evidence that the residual discrepancies in the sequence effect on the BI and BII populations resulted from insufficient sampling. The BII percentages were found converged well before the half microsecond timescale under monitored MD length increase. Since the BI↔BII exchange occurs in the pico to nanosecond time range [19, 46], which is short compared to current simulation times, one does not expect that the BI↔BII equilibrium distribution will improve by increasing the sampling time. Instead, progress is likely to require further refinements of the potentials. Considering the charged nature of the phosphate groups, it is possible that polarisable DNA force-fields will be required to reach a more satisfactory treatment of the sequence-dependent DNA properties [3, 61].

Despite some limitations, the present MDs are very helpful to scrutinize the DNA backbone dynamics, especially the behavior of facing phosphate groups within complementary dinucleotides. In that respect, Parmbsc0_{ε_{OL}} or CHARMM36 yielded similar features despite strong differences in their conception and parametrizations. This convergence strengthens the results. First the simulations indicate that concomitant BI ↔ BII transitions on two facing phosphates are much rarer than transitions involving only one phosphate. Second, statistical analysis of the simulations established that the conformational states of the two individual phosphates within a complementary dinucleotide were independent of each other. As a consequence, the BI•BI, BI•BII|BII•BI and BII•BII populations can be assessed from the individual BII percentages inferred from δPs, using straightforward equations. Importantly, this approach reveals that there is a sizable number of steps where BII-containing states dominate. Thus, more than one fourth of the 72 complementary dinucleotides spend more time in BI•BII|

BII•BI and BII•BII than in BI•BI; all the complementary dinucleotides explore BI•BII|BII•BI in addition to BI•BI, with various populations of BI•BII|BII•BI; however, BII•BII is more restricted, apparently only significantly populated in a few types of BII-rich steps.

Since the behavior of facing phosphates was uncorrelated in all the 36 complementary steps studied here, one can reasonably infer that this is a general property of any B-DNA. Thus, according to the general and predictable sequence effect on experimental BII propensities [22, 23], the BII-containing combinations (BI•BII|BII•BI and BII•BII) are expected to be largely represented or even statistically dominant in CpG•CpG, CpA•TpG, GpC•GpC and GpG•CpC. The steps less propitious to BII, GpA•TpC, ApN•NpT (N: any base) and TpA•TpA, favor BI•BI but they also present modest fractions of BI•BII|BII•BI.

Such findings are of fundamental importance because of the strong couplings between these fine-grained backbone states and the inter base pair parameters of slide, roll and twist, consistent with initial observations on X-ray structures [11, 12]. Such couplings are not only confirmed here, but further characterized in solution for a broader range of steps, offering a unifying theme underpinning the intrinsic mechanics of B-DNA. Given the recurrent occurrence of BII-containing combinations, it follows that the accessible conformational landscape of most complementary dinucleotides extends into a region characterized by positive slide, negative roll and high twist (“BII profile”).

This enhanced intrinsic malleability is relevant to the reading of DNA by proteins, since it increases the repertoire of states which may be critical to initiate selective recognition by facilitating local, structural DNA adjustments upon protein binding. The implication of BII-rich steps in indirect readout mechanisms, *via* their ability to modulate the DNA shape, has been previously highlighted [5, 38, 62, 63]. In addition, the present work touched upon the counterintuitive example of the BI-rich (positive rolls) TTAAA segment in Oligo 3, which nevertheless also accesses negative rolls (BII•BI|BI•BII) in solution, reminiscent of the pattern of negative rolls observed in its nucleosome-bound form. So, the energetic penalty induced by the DNA deformation upon protein binding could be less than expected in many cases, especially when BII-like features are involved for the structural fit between the partners. Thus, the present characterization of free DNA is conceptually relevant to a deeper understanding of the selective recognition of DNA. The investigated force-fields Parmbsc0_{OL1} or CHARMM36 may also prove advantageous when simulating such events.

MATERIAL AND METHODS

DNA sequences

Four oligodeoxyribonucleotides of 12 base pairs (bp) (sequences in Table 1) were studied by NMR and MD simulations. These sequences, placed end to end after discarding the terminal base pairs, recompose a continuous 39 bp segment corresponding to the 5' part of the non-palindromic sequence 601, selected from SELEX experiments for its very high-affinity for association with the histone octamers [44].

BII propensities from NMR

Sample preparation and NMR spectroscopy protocols were reported in a previous study [23]. All the NMR data are available in the Biological Magnetic Resonance Bank, entry 19222.

BII percentages (BII%) of the phosphate linkages along the four dodecamers were inferred from the phosphate chemical shifts (δP s, referenced to trimethyl phosphate) collected at 30°, using the equation $BII(\%) = 143 \delta P + 621$ [20]. This equation is based on an empirical procedure that assumes the same δP s for purely BI or BII states of every dinucleotide, which is unlikely to be strictly correct [48]. Although previous studies showed that it is a reasonable approximation [23, 49], we decided to allow a large tolerance of $\pm 10\%$ on the BII percentages inferred from the experimental δP s to take into account uncertainty on the translation procedure.

MDs with Parmbsc0_{εζOLI} and CHARMM36 force fields

MD simulations were performed with the Parmbsc0_{εζOLI} force-field [9] using the AMBER 14 program [64], or the CHARMM36 force-field [7] with program NAMD [65]. Parmbsc0_{εζOLI} and CHARMM36 simulations were carried out following protocols as comparable as possible. Yet, with Parmbsc0_{εζOLI} and CHARMM36, we used the counterion parameters classically associated to the Amber [66] and CHARMM [67] force-fields, respectively.

Parmbsc0_{εζOLI} and CHARMM36 simulations were performed at constant temperature (300K) and pressure (1bar) using the Berendsen algorithm [68]. The integration time-step was 2fs and covalent bonds involving hydrogen were constrained using SHAKE [69]. The non-bonded pair-list was updated heuristically. Long-range electrostatic interactions were treated using the particle mesh Ewald (PME) approach [70]. Non-bonded interactions were treated with a 9Å direct space cut-off in AMBER and with a force-shift function from 10 to 12 Å [71] with CHARMM36. In AMBER, the centre-of-mass motion was removed every 10ps.

With both Parmbsc0_{εζOLI} and CHARMM36, each dodecamer in initial standard B-DNA conformation was neutralized with 22 Na⁺ ions (minimal salt condition, ~50 mM Na⁺), in explicit TIP3P water molecules [72]; the primary boxes were truncated octahedrons with solvent extending 15Å around the DNA. The water molecules and counterions were energy-minimized and equilibrated at 100K around the constrained DNA for 100ps in the NVT ensemble; the entire system was then heated from 100 to 300K in 10ps by 5K increments with harmonic positional restraints of 5.0 kcal/mol/Å² on the DNA atoms. The molecular dynamics simulations were continued in NPT, without notable change in volume. The positional restraints were gradually removed over 250ps and followed by the production phase. During the simulations, distance restraints were applied between base atoms of the first and last base pairs of each dodecamers, to prevent their opening. No restraint was applied on any of the internal nucleotides. The application of restraints on the terminal base pairs is justified in the next section, which highlights the benefits of conducting DNA simulations work alongside experimental characterization. MD snapshots were saved every 1 ps.

Restrained base pairing on the first and last base pairs

During the simulations with Parmbsc0_{ε_{OLI}} and CHARMM36, distance restraints were applied to maintain the Watson-Crick base-pairing in the first and last base pairs of each dodecamers, to prevent their opening. These restraints were applied on the terminal base pairs between base atoms involved in Watson-Crick hydrogen-bonding ($\text{Distance}_{\text{donor/acceptor}} = 2.9 \pm 0.2 \text{ \AA}$) via a parabolic potential with a force-constant of $10 \text{ kcal/mol/\AA}^2$. The application of these restraints was motivated by the behavior of terminal base-pairs in unrestrained simulations, which are not presented in details here. We only give a summary of the unrestrained terminal base-pairs simulations compared to relevant NMR data, to justify the application of restraints in the presented MDs.

In the unrestrained simulations with Parmbsc0_{ε_{OLI}}, the first ($\text{N}_1:\text{N}_{24}$, N for any base type) and last ($\text{N}_{12}:\text{N}_{13}$) base pairs were generally open. These terminal bases, once extruded, got involved in various structural patterns that persisted during several hundreds of nanoseconds. In the most prevalent conformations, these bases interact with the penultimate phosphate group, insert into the minor groove or mispair with an antepenultimate base. These conformations impact some χ angles, are associated with unusual backbone dihedrals in N_1pN_2 , $\text{N}_{11}\text{pN}_{12}$, $\text{N}_{13}\text{pN}_{14}$ or $\text{N}_{23}\text{pN}_{24}$, and break the stacking with the 3' or 5' neighbors (N_2 , N_{11} , N_{14} or N_{23}). With CHARMM36, the first two base pairs opened after a few nanoseconds and, as in Parmbsc0_{ε_{OLI}} MDs, adopted multiple non-canonical structures. In these unrestrained MDs, base pair opening only occurred at the termini of the dodecamers and did not propagate further. Such behavior is not specific to our simulations since it was previously described for MDs with Parmbsc0 and Parmbsc0_{OLI} [1, 9, 73] or CHARMM36 [1], which used DNA sequences and simulation protocols different from the ones used here.

The re-orientation of the terminal bases towards the internal double stranded part of the DNA are not supported by the NMR data collected at 20 and 30°C on the four dodecamers. In one-dimensional ^1H spectra at 30° (303K), the imino proton resonances are lost in the terminal base pairs while they are clearly observable in all the internal base pairs (from $\text{N}_2:\text{N}_{23}$ to $\text{N}_{11}:\text{N}_{14}$). This excludes long-lived disruption of the Watson-Crick hydrogen bonds in the penultimate base pairs (MDs with CHARMM36) or mispairing between a terminal base and an antepenultimate base (MDs with Parmbsc0_{ε_{OLI}}). The glycosidic bonds of the terminal nucleotides, probed by the intranucleotide distances $\text{H1}'\text{-H6}/8$, adopt the *anti* conformation. Furthermore, the numerous sequential NOEs between the penultimate and antepenultimate residues (N_2pN_3 , $\text{N}_{10}\text{pN}_{11}$, $\text{N}_{14}\text{pN}_{15}$ or $\text{N}_{22}\text{pN}_{23}$) do not support extensive break of their stacking or abnormal structural features. NMR measurements also give information about the terminal steps, N_1pN_2 , $\text{N}_{11}\text{pN}_{12}$, $\text{N}_{13}\text{pN}_{14}$ and $\text{N}_{23}\text{pN}_{24}$. The corresponding ^{31}P chemical shifts are in the range of the internal phosphates. Intense, well defined $^{31}\text{P}\text{-}^1\text{H4}'$ couplings testify that the 3' terminal phosphate groups conform to usual backbone conformation, since these couplings are observable only when $\alpha/\beta/\gamma$ are in *g*-/*trans*/*g*+ [19, 74], the typical conformation of B-DNA. Finally, sequential NOE connectivities, clearly visible in all the terminal steps, imply that the fraying events do not generate large distance between open terminal bases and the penultimate residues.

In agreement with a detailed study of this topic [73], our NMR data indicate that current force fields do not yet provide a satisfactory description of the fraying of the terminal base pairs. In addition, the disruption of the structure and dynamics of the terminal regions in our unrestrained MDs induced convergence issues on neighboring steps. Restraining the first and last base pairs enabled to overcome these shortcomings. Reassuringly, the structural characteristics

calculated on the central part of the dodecamers are virtually identical in restrained and unrestrained Parmbsc0_{ε_{OLI}} and CHARMM36 MDs, indicating that restraining the terminal base pairs does not unduly alter the behavior of the internal base-pairs.

Structural descriptors

The phosphate group linkages were characterized by torsion angles ϵ , ζ , α , β and γ following the conventional threefold staggered torsional pattern: *gauche plus* ($60\pm 40^\circ$), *trans* ($180\pm 40^\circ$) and *gauche minus* ($300\pm 40^\circ$). The sugar ring conformations were categorized according to their pseudorotation phase angle: *north* (300 to 50°), *east* (50 to 120°) and *south* (120 to 220°).

DNA structures were analyzed with Curves5 [75] and 3DNA [76]. Both programs produced almost identical helical parameter values. The inter base-pair parameters presented here for complementary dinucleotides NpN•NpN are those from Curves5. Only the 10 central base-pairs of each dodecamer were analyzed.

Acknowledgments

The authors thank Dr Olivier Mauffret (LBPA, CNRS / ENS de Cachan) for helpful advice. The simulations were carried out on the GENCI-CEA platform.

References

- Galindo-Murillo R, Roe DR, Cheatham TE, 3rd. Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). *Biochim Biophys Acta*. 2015;1850(5):1041-58. Epub 2014/09/16. doi: 10.1016/j.bbagen.2014.09.007. PubMed PMID: 25219455; PubMed Central PMCID: PMC4339415.
- Perez A, Luque FJ, Orozco M. Frontiers in Molecular Dynamics Simulations of DNA. *Accounts of Chemical Research*. 2012;45(2):196-205. doi: 10.1021/ar2001217.
- Vanommeslaeghe K, MacKerell AD, Jr. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim Biophys Acta*. 2015;1850(5):861-71. Epub 2014/08/26. doi: 10.1016/j.bbagen.2014.08.004. PubMed PMID: 25149274.
- Bosch D, Foloppe N, Pastor N, Pardo L, Campillo M. Calibrating nucleic acids torsional energetics in force-field: insights from model compounds. *Journal of Molecular Structure: THEOCHEM*. 2001;537(1-3):283-305. doi: [http://dx.doi.org/10.1016/S0166-1280\(00\)00685-0](http://dx.doi.org/10.1016/S0166-1280(00)00685-0).
- Heddi B, Foloppe N, Oguey C, Hartmann B. Importance of accurate DNA structures in solution: the Jun-Fos model. *Journal of molecular biology*. 2008;382(4):956-70. Epub 2008/08/06. doi: 10.1016/j.jmb.2008.07.047. PubMed PMID: 18680751.
- Zuo X, Cui G, Merz KM, Jr., Zhang L, Lewis FD, Tiede DM. X-ray diffraction "fingerprinting" of DNA structure in solution for quantitative evaluation of molecular dynamics simulation. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(10):3534-9. Epub 2006/03/01. doi: 10.1073/pnas.0600022103. PubMed PMID: 16505363.
- Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, Mackerell AD, Jr. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J Chem Theory Comput*. 2012;8(1):348-62. Epub 2012/03/01. doi: 10.1021/ct200723y. PubMed PMID: 22368531; PubMed Central PMCID: PMC3285246.
- Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, 3rd, Laughton CA, et al. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophysical journal*. 2007;92(11):3817-29. Epub 2007/03/14. doi: 10.1529/biophysj.106.097782. PubMed PMID: 17351000; PubMed Central PMCID: PMC1868997.
- Zgarbova M, Luque FJ, Sponer J, Cheatham TE, 3rd, Otyepka M, Jurecka P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J Chem Theory Comput*. 2013;9(5):2339-54. Epub 2013/09/24. doi: 10.1021/ct400154j. PubMed PMID: 24058302; PubMed Central PMCID: PMC3775469.
- Berman HM. Crystal studies of B-DNA: the answers and the questions. *Biopolymers*. 1997;44(1):23-44. Epub 1997/01/01. doi: 10.1002/(SICI)1097-0282(1997)44:1<23::AID-BIP3>3.0.CO;2-1. PubMed PMID: 9097732.
- Djuranovic D, Hartmann B. Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *Journal of biomolecular structure & dynamics*. 2003;20(6):771-88. PubMed PMID: 12744707.

12. Djuranovic D, Hartmann B. DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers*. 2004;73(3):356-68. PubMed PMID: 14755572.
13. Foloppe N, MacKerell AD. Contribution of the Phosphodiester Backbone and Glycosyl Linkage Intrinsic Torsional Energetics to DNA Structure and Dynamics. *The Journal of Physical Chemistry B*. 1999;103(49):10955-64. doi: 10.1021/jp992716q.
14. Schneider B, Neidle S, Berman HM. Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers*. 1997;42(1):113-24. Epub 1997/01/01. PubMed PMID: 19350745.
15. Varnai P, Djuranovic D, Lavery R, Hartmann B. Alpha/gamma transitions in the B-DNA backbone. *Nucleic acids research*. 2002;30(24):5398-406. PubMed PMID: 12490708.
16. Fratini AV, Kopka ML, Drew HR, Dickerson RE. Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J Biol Chem*. 1982;257(24):14686-707. Epub 1982/12/25. PubMed PMID: 7174662.
17. Gorenstein DG. *Phosphorus-31 NMR: Principles and Applications*: Academic Press, New York; 1984.
18. Oguey C, Foloppe N, Hartmann B. Understanding the sequence-dependence of DNA groove dimensions: implications for DNA interactions. *PLoS One*. 2010;5(12):e15931. Epub 2011/01/07. doi: 10.1371/journal.pone.0015931. PubMed PMID: 21209967; PubMed Central PMCID: PMC3012109.
19. Gorenstein DG. ³¹P NMR of DNA. *Methods Enzymol*. 1992;211:254-86. PubMed PMID: 1406310.
20. Heddi B, Foloppe N, Bouchemal N, Hantz E, Hartmann B. Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition. *Journal of the American Chemical Society*. 2006;128(28):9170-7. Epub 2006/07/13. doi: 10.1021/ja061686j. PubMed PMID: 16834390.
21. Tian Y, Kayatta M, Shultis K, Gonzalez A, Mueller LJ, Hatcher ME. (31)P NMR Investigation of Backbone Dynamics in DNA Binding Sites (dagger). *J Phys Chem B*. 2009;113(9):2596-603. Epub 2008/08/23. doi: 10.1021/jp711203m. PubMed PMID: 18717548; PubMed Central PMCID: PMC2711773.
22. Heddi B, Oguey C, Lavelle C, Foloppe N, Hartmann B. Intrinsic flexibility of B-DNA: the experimental TRX scale. *Nucleic acids research*. 2010;38(3):1034-47. Epub 2009/11/19. doi: 10.1093/nar/gkp962. PubMed PMID: 19920127; PubMed Central PMCID: PMC2817485.
23. Xu X, Ben Imeddourene A, Zargarian L, Foloppe N, Mauffret O, Hartmann B. NMR studies of DNA support the role of pre-existing minor groove variations in nucleosome indirect readout. *Biochemistry*. 2014;53(35):5601-12. Epub 2014/08/08. doi: 10.1021/bi500504y. PubMed PMID: 25102280.
24. Cheatham TE, 3rd, Cieplak P, Kollman PA. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *Journal of biomolecular structure & dynamics*. 1999;16(4):845-62. Epub 1999/04/27. PubMed PMID: 10217454.
25. Wang J, Cieplak P, Kollman P. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of computational chemistry*. 2000;21(12):1049-74.

26. Varnai P, Zakrzewska K. DNA and its counterions: a molecular dynamics study. *Nucleic acids research*. 2004;32(14):4269-80. Epub 2004/08/12. doi: 10.1093/nar/gkh765. PubMed PMID: 15304564; PubMed Central PMCID: PMC514386.
27. Lefebvre A, Mauffret O, Hartmann B, Lescot E, Fermandjian S. Structural behavior of the CpG step in two related oligonucleotides reflects its malleability in solution. *Biochemistry*. 1995;34(37):12019-28. PubMed PMID: 7547940.
28. Lefebvre A, Mauffret O, Lescot E, Hartmann B, Fermandjian S. Solution structure of the CpG containing d(CTTCGAAG)₂ oligonucleotide: NMR data and energy calculations are compatible with a BI/BII equilibrium at CpG. *Biochemistry*. 1996;35(38):12560-9. PubMed PMID: 8823193.
29. Dans PD, Faustino I, Battistini F, Zakrzewska K, Lavery R, Orozco M. Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA. *Nucleic acids research*. 2014;42(18):11304-20. Epub 2014/09/17. doi: 10.1093/nar/gku809. PubMed PMID: 25223784.
30. Heddi B, Foloppe N, Hantz E, Hartmann B. The DNA structure responds differently to physiological concentrations of K(+) or Na(+). *Journal of molecular biology*. 2007;368(5):1403-11. PubMed PMID: 17395202.
31. Foloppe N, MacKerell AD. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry*. 2000;21(2):86 - 104.
32. Mackerell AD, Jr. Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem*. 2004;25(13):1584-604. Epub 2004/07/21. doi: 10.1002/jcc.20082. PubMed PMID: 15264253.
33. Perez A, Lankas F, Luque FJ, Orozco M. Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic acids research*. 2008;36(7):2379-94. PubMed PMID: 18299282.
34. Foloppe N, Gueroult M, Hartmann B. Simulating DNA by molecular dynamics: aims, methods, and validation. *Methods Mol Biol*. 2013;924:445-68. Epub 2012/10/05. doi: 10.1007/978-1-62703-017-5_17. PubMed PMID: 23034759.
35. Drsata T, Perez A, Orozco M, Morozov AV, Sponer J, Lankas F. Structure, Stiffness and Substates of the Dickerson-Drew Dodecamer. *J Chem Theory Comput*. 2013;9(1):707-21. Epub 2013/08/27. doi: 10.1021/ct300671y. PubMed PMID: 23976886; PubMed Central PMCID: PMC3748671.
36. Pasi M, Maddocks JH, Beveridge D, Bishop TC, Case DA, Cheatham T, 3rd, et al. muABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic acids research*. 2014;42(19):12272-83. Epub 2014/09/28. doi: 10.1093/nar/gku855. PubMed PMID: 25260586; PubMed Central PMCID: PMC4231739.
37. Botuyan MV, Keire DA, Kroen C, Gorenstein DG. 31P nuclear magnetic resonance spectra and dissociation constants of lac repressor headpiece.duplex operator complexes: the importance of phosphate backbone flexibility in protein.DNA recognition. *Biochemistry*. 1993;32(27):6863-74. Epub 1993/07/13. PubMed PMID: 8334119.
38. Heddi B, Abi-Ghanem J, Lavigne M, Hartmann B. Sequence-dependent DNA flexibility mediates DNase I cleavage. *Journal of molecular biology*. 2010;395(1):123-33. Epub 2009/10/24. doi: 10.1016/j.jmb.2009.10.023. PubMed PMID: 19850052.

39. Karslake C, Botuyan MV, Gorenstein DG. 31P NMR spectra of oligodeoxyribonucleotide duplex lac operator-repressor headpiece complexes: importance of phosphate ester backbone flexibility in protein-DNA recognition. *Biochemistry*. 1992;31(6):1849-58. Epub 1992/02/18. PubMed PMID: 1737038.
40. Tisne C, Hantz E, Hartmann B, Delepierre M. Solution structure of a non-palindromic 16 base-pair DNA related to the HIV-1 kappa B site: evidence for BI-BII equilibrium inducing a global dynamic curvature of the duplex. *Journal of molecular biology*. 1998;279(1):127-42. PubMed PMID: 9636705.
41. Tisne C, Hartmann B, Delepierre M. NF-kappa B binding mechanism: a nuclear magnetic resonance and modeling study of a GGG --> CTC mutation. *Biochemistry*. 1999;38(13):3883-94. Epub 1999/04/09. doi: 10.1021/bi982402d. PubMed PMID: 10194299.
42. Wecker K, Bonnet MC, Meurs EF, Delepierre M. The role of the phosphorus BI-BII transition in protein-DNA recognition: the NF-kappaB complex. *Nucleic acids research*. 2002;30(20):4452-9. Epub 2002/10/18. PubMed PMID: 12384592.
43. Nikolova EN, Bascom GD, Andricioaei I, Al-Hashimi HM. Probing sequence-specific DNA flexibility in a-tracts and pyrimidine-purine steps by nuclear magnetic resonance (13)C relaxation and molecular dynamics simulations. *Biochemistry*. 2012;51(43):8654-64. Epub 2012/10/06. doi: 10.1021/bi3009517. PubMed PMID: 23035755; PubMed Central PMCID: PMC3676944.
44. Thastrom A, Bingham LM, Widom J. Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. *Journal of molecular biology*. 2004;338(4):695-709. PubMed PMID: 15099738.
45. Galindo-Murillo R, Roe DR, Cheatham TE, 3rd. On the absence of intrahelical DNA dynamics on the ms to ms timescale. *Nat Commun*. 2014;5:5152. Epub 2014/10/30. doi: 10.1038/ncomms6152. PubMed PMID: 25351257; PubMed Central PMCID: PMC4215645.
46. Gorenstein DG. Conformation and Dynamics of DNA and Protein-DNA Complexes by 31P NMR. *Chemical Review*. 1994;94(5):1315-38.
47. Isaacs RJ, Spielmann HP. NMR evidence for mechanical coupling of phosphate B(I)-B(II) transitions with deoxyribose conformational exchange in DNA. *Journal of molecular biology*. 2001;311(1):149-60. Epub 2001/07/27. doi: 10.1006/jmbi.2001.4855. PubMed PMID: 11469864.
48. Precechtelova J, Munzarova ML, Vaara J, Novotny J, Dracinski M, Sklenar V. Toward Reproducing Sequence Trends in Phosphorus Chemical Shifts for Nucleic Acids by MD/DFT Calculations. *Journal of Chemical Theory and Computation*. 2013;9(3):1641-56. doi: 10.1021/ct300488y.
49. Schwieters CD, Clore GM. A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry*. 2007;46(5):1152-66. Epub 2007/01/31. doi: 10.1021/bi061943x. PubMed PMID: 17260945.
50. Mackerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*. 2004;25(11):1400-15. doi: 10.1002/jcc.20065.

51. Hartmann B, Piazzola D, Lavery R. BI-BII transitions in B-DNA. *Nucleic acids research*. 1993;21(3):561-8. Epub 1993/02/11. PubMed PMID: 8441668.
52. Srinivasan AR, Olson WK. Nucleic acid model building: the multiple backbone solutions associated with a given base morphology. *Journal of biomolecular structure & dynamics*. 1987;4(6):895-938. Epub 1987/06/01. PubMed PMID: 3270539.
53. van Dam L, Levitt MH. BII nucleotides in the B and C forms of natural-sequence polymeric DNA: A new model for the C form of DNA. *Journal of molecular biology*. 2000;304(4):541-61. Epub 2000/12/02. doi: 10.1006/jmbi.2000.4194. PubMed PMID: 11099379.
54. Winger RH, Liedl KR, Pichler A, Hallbrucker A, Mayer E. Helix morphology changes in B-DNA induced by spontaneous B(I) \rightleftharpoons B(II) substrate interconversion. *Journal of biomolecular structure & dynamics*. 1999;17(2):223-35. Epub 1999/11/24. PubMed PMID: 10563572.
55. Hahn M, Heinemann U. DNA helix structure and refinement algorithm: comparison of models for d(CCAGGCm5CTGG) derived from NUCLSQ, TNT and X-PLOR. *Acta crystallographica*. 1993;49(Pt 5):468-77. Epub 1993/09/01. doi: 10.1107/S0907444993004858. PubMed PMID: 15299506.
56. Chua EY, Vasudevan D, Davey GE, Wu B, Davey CA. The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic acids research*. 2012;40(13):6338-52. Epub 2012/03/29. doi: 10.1093/nar/gks261. PubMed PMID: 22453276; PubMed Central PMCID: PMC3401446.
57. Makde RD, England JR, Yennawar HP, Tan S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature*. 2010;467(7315):562-6. Epub 2010/08/27. doi: 10.1038/nature09321. PubMed PMID: 20739938; PubMed Central PMCID: PMC3168546.
58. Ong MS, Richmond TJ, Davey CA. DNA stretching and extreme kinking in the nucleosome core. *Journal of molecular biology*. 2007;368(4):1067-74. Epub 2007/03/24. doi: 10.1016/j.jmb.2007.02.062. PubMed PMID: 17379244.
59. Vasudevan D, Chua EY, Davey CA. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *Journal of molecular biology*. 2010;403(1):1-10. Epub 2010/08/31. doi: 10.1016/j.jmb.2010.08.039. PubMed PMID: 20800598.
60. Wu B, Mohideen K, Vasudevan D, Davey CA. Structural insight into the sequence dependence of nucleosome positioning. *Structure*. 2010;18(4):528-36. Epub 2010/04/20. doi: 10.1016/j.str.2010.01.015. PubMed PMID: 20399189.
61. Savelyev A, MacKerell AD, Jr. All-atom polarizable force field for DNA based on the classical Drude oscillator model. *J Comput Chem*. 2014;35(16):1219-39. Epub 2014/04/23. doi: 10.1002/jcc.23611. PubMed PMID: 24752978; PubMed Central PMCID: PMC4075971.
62. Djuranovic D, Hartmann B. Molecular dynamics studies on free and bound targets of the bovine papillomavirus type I e2 protein: the protein binding effect on DNA and the recognition mechanism. *Biophysical journal*. 2005;89(4):2542-51. PubMed PMID: 16055534.
63. Tisne C, Delepierre M, Hartmann B. How NF-kappaB can be attracted by its cognate DNA. *Journal of molecular biology*. 1999;293(1):139-50. PubMed PMID: 10512722.

64. Case DA, Darden TA, Cheatham I, T.E., Simmerling CL, Wang J, Duke RE, et al. AMBER 9. University of California, San Francisco. 2006.
65. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*. 2005;26(16):1781-802. doi: 10.1002/jcc.20289.
66. Aqvist J. Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J Phys Chem*. 1990;94:8021–24.
67. Venable RM, Luo Y, Gawrisch K, Roux B, Pastor RW. Simulations of anionic lipid membranes: development of interaction-specific ion parameters and validation using NMR data. *J Phys Chem B*. 2013;117(35):10183-92. Epub 2013/08/09. doi: 10.1021/jp401512z. PubMed PMID: 23924441; PubMed Central PMCID: PMC3813009.
68. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*. 1984;81(8):3684-90.
69. van Gunsteren WF, Berendsen HJC. Algorithms for macromolecular dynamics and constraint dynamics. *Molecular Physics*. 1977;34(5):1311-27.
70. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys*. 1993;98(12):10089-92.
71. Steinbach PJ, Brooks BR. New spherical-cutoff methods for long-range forces in macromolecular simulation. *Journal of Computational Chemistry*. 1994;15(7):667-83. doi: 10.1002/jcc.540150702.
72. Jorgensen WL, Chandrasekhar J, Madura JD. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926-35.
73. Zgarbova M, Otyepka M, Sponer J, Lankas F, Jurecka P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *Journal of Chemical Theory and Computation*. 2014;10(8):3177-89. doi: 10.1021/ct500120v.
74. Zhu L, Chou SH, Reid BR. A single G-to-C change causes human centromere TGGAA repeats to fold back into hairpins. *Proceedings of the National Academy of Sciences of the United States of America*. 1996;93(22):12159-64. Epub 1996/10/29. PubMed PMID: 8901550; PubMed Central PMCID: PMC37960.
75. Lavery R, Sklenar H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *Journal of biomolecular structure & dynamics*. 1988;6(1):63-91. Epub 1988/08/01. PubMed PMID: 2482765.
76. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic acids research*. 2003;31(17):5108-21. Epub 2003/08/22. PubMed PMID: 12930962; PubMed Central PMCID: PMC212791.

SUPPLEMENTARY DATA

FIGURES

S1 Fig. RMSD along the DNA trajectories with Parmbsc0_{εζOLI} and CHARMM36.

S2 Fig. Convergence of BII percentages in 450ns and 1μs MD simulations.

S3 Fig. (ϵ - ζ) distribution in X-ray structures.

S4 Fig. Influence of the force-field on the sugar puckers during the MD simulation of Oligo 4.

S5 Fig. Comparison between simulated and experimental BII percentages.

S6 Fig. Slide, Roll and Twist values associated to conformational combinations of facing phosphate linkages in representative BII-rich steps, from C-MDs and P-MDs.

S7 Fig. Comparison between the conformational combinations of facing phosphate linkages extracted from C-MDs or based on NMR data.

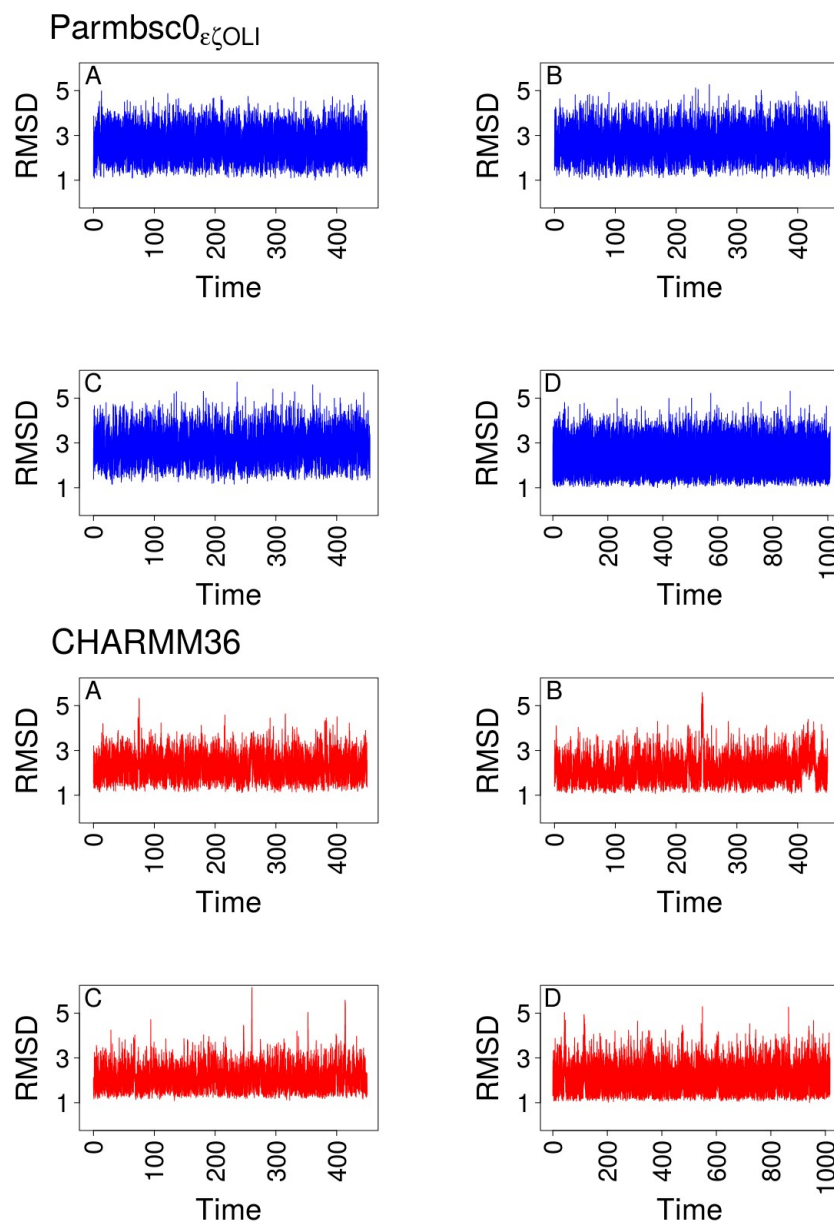
TABLES

S1 Table. BII percentages from NMR and simulations with Parmbsc0_{εζOLI} and CHARMM36.

S2 Table. Populations of the three conformational combinations of facing phosphate groups based on experimental data.

S3 Table. Values of inter base pair parameters according to the conformational combinations of facing phosphate groups.

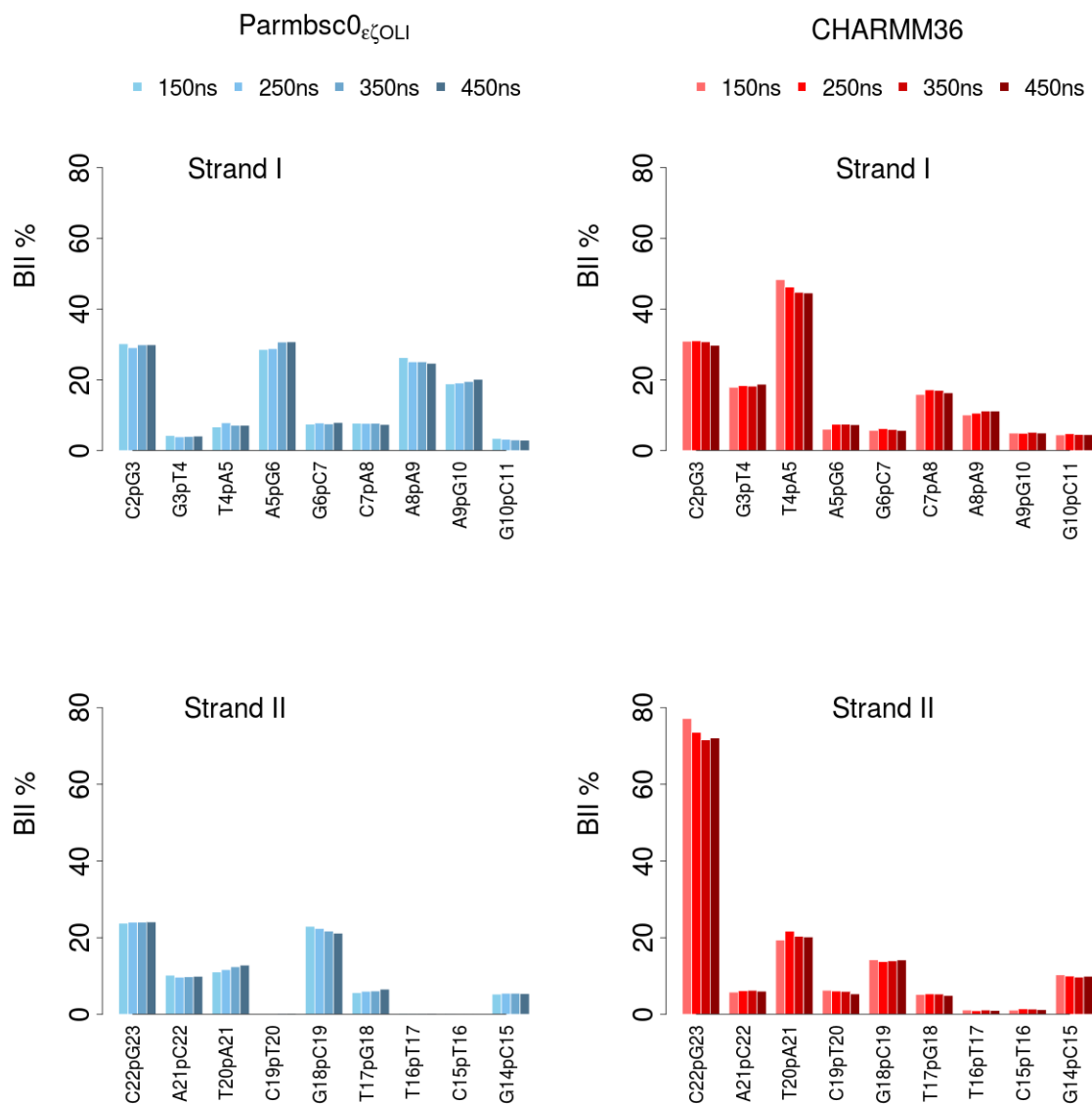
S1 Fig. RMSD along the trajectories with Parmbsc0 _{$\epsilon\zeta$ OLI} and CHARMM36.



The Root Mean Square Deviations (RMSD, Å) are plotted in function of time (ns) for four DNA dodecamers (A: Oligo 1; B: Oligo 2; C: Oligo 3; D: Oligo 4) simulated with either Parmbsc0 _{$\epsilon\zeta$ OLI} (top panels, in blue) or CHARMM36 (bottom panels, in red) force fields. RMSDs were calculated on the heavy atoms of the simulated structures -discarding the first and last base-pairs- and reference canonical B-DNAs generated by AMBER.

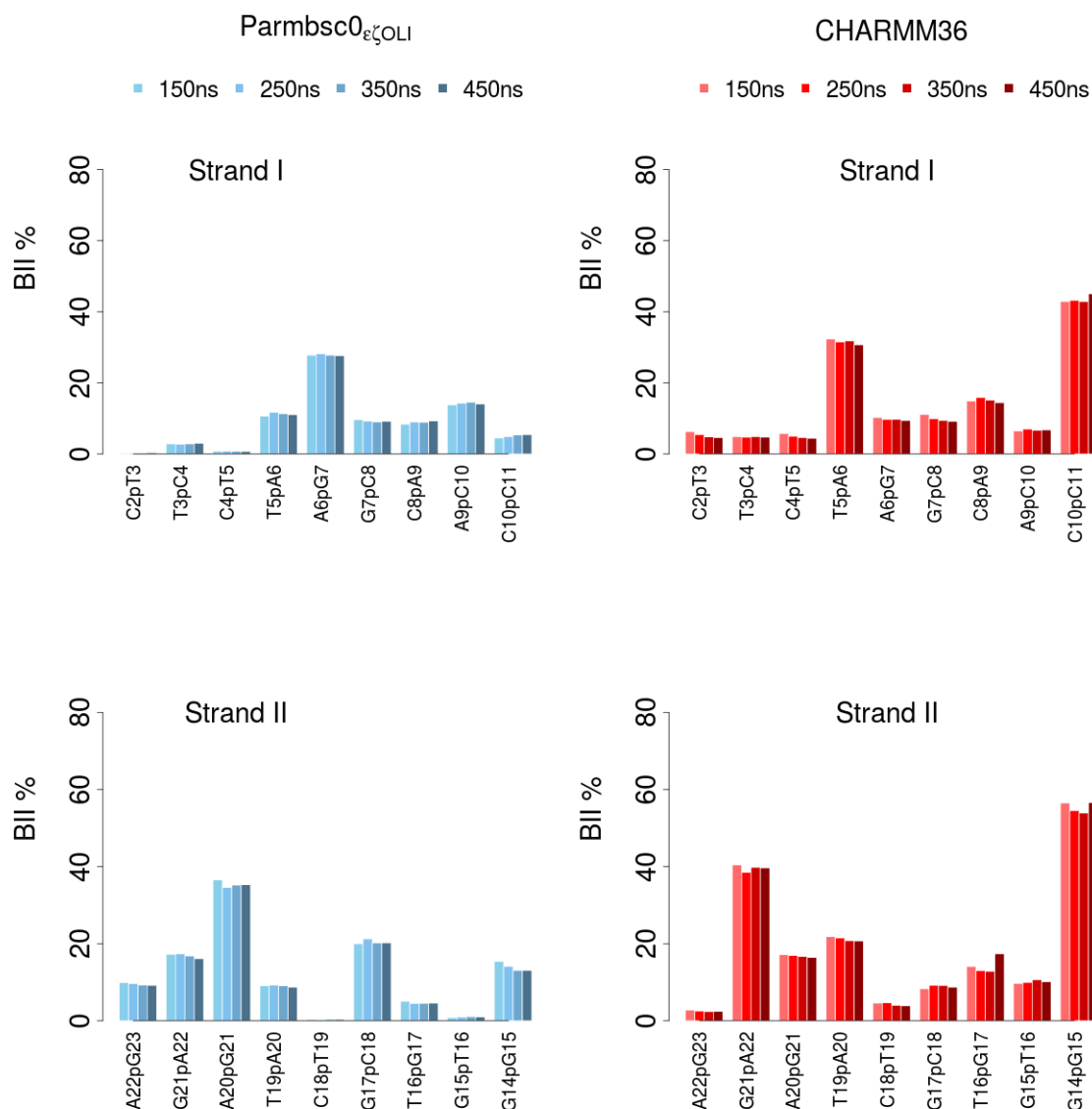
S2 Fig. Convergence of BII percentages in 450ns and 1 μ s MD simulations.

Oligo 1



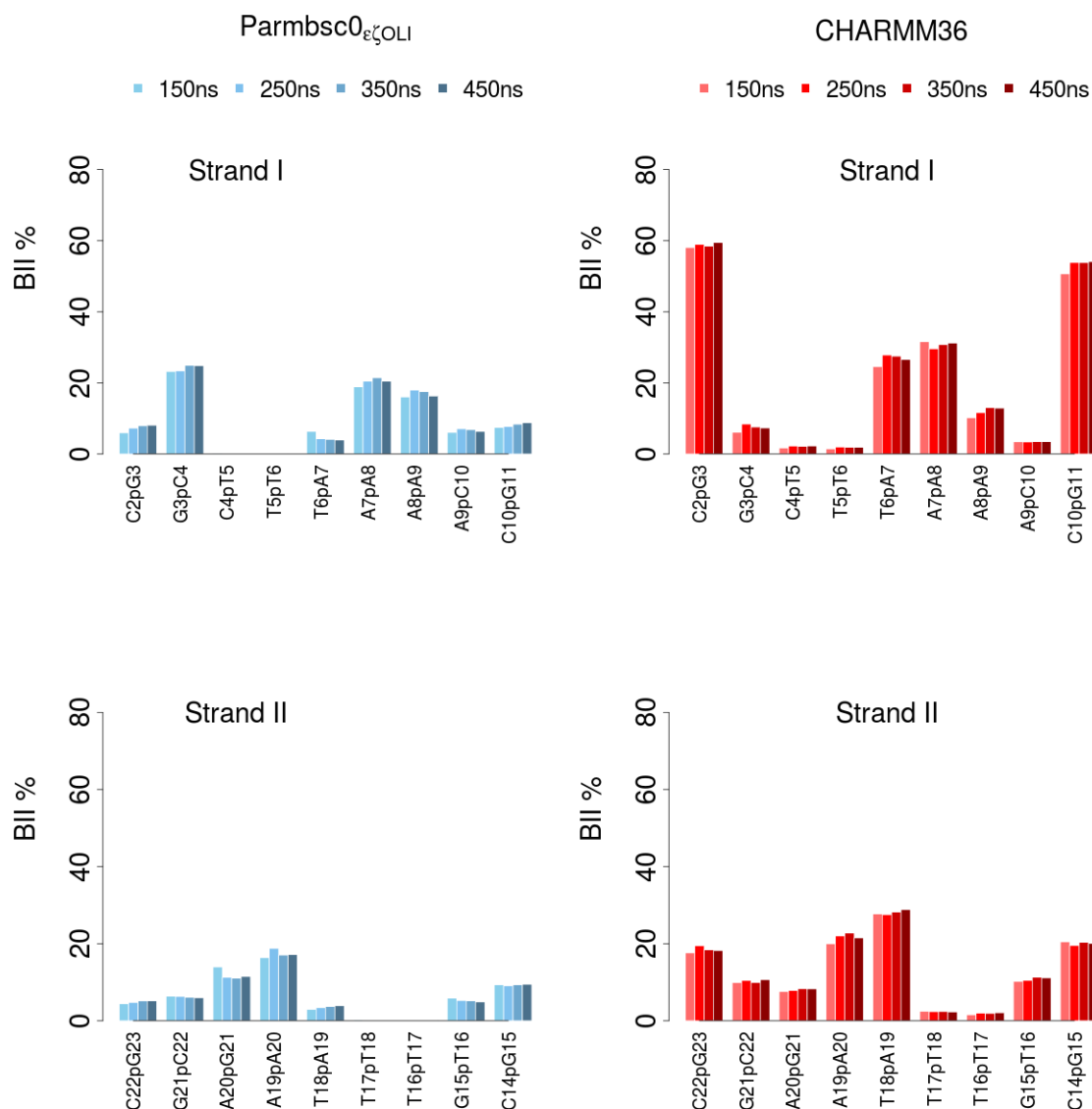
Cumulative BII percentages (BII%) of the dinucleotide steps were extracted from trajectories carried out with either Parmbsc0_{εζOLI} (left, blue) or CHARMM36 (right, red) force fields. BII% were calculated for increasing length of MDs. The BII% were clearly not stabilized during the first 50ns, which are not considered in this analysis. To help identification of the facing steps, sequence in Strand 1 is shown from 5' to 3', while sequence in Strand 2 is shown from 3' to 5'.

Oligo 2



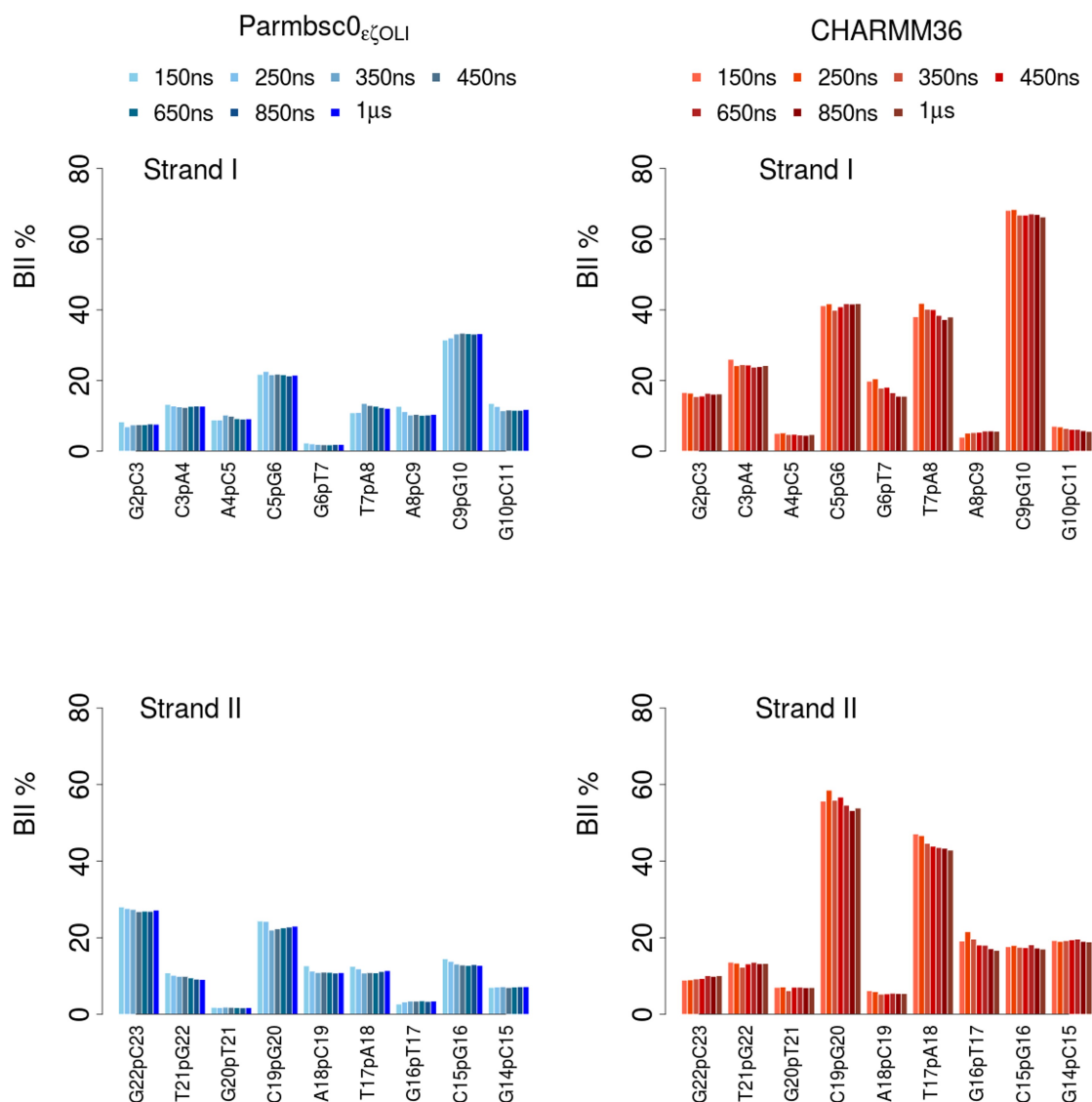
Cumulative BII percentages (BII%) of the dinucleotide steps were extracted from trajectories carried out with either Parmbsc0 $\epsilon\zeta$ OLI (left, blue) or CHARMM36 (right, red) force fields. BII% were calculated for increasing length of MDs. The BII% were clearly not stabilized during the first 50ns, which are not considered in this analysis. To help identification of the facing steps, sequence in Strand 1 is shown from 5' to 3', while sequence in Strand 2 is shown from 3' to 5'.

Oligo 3



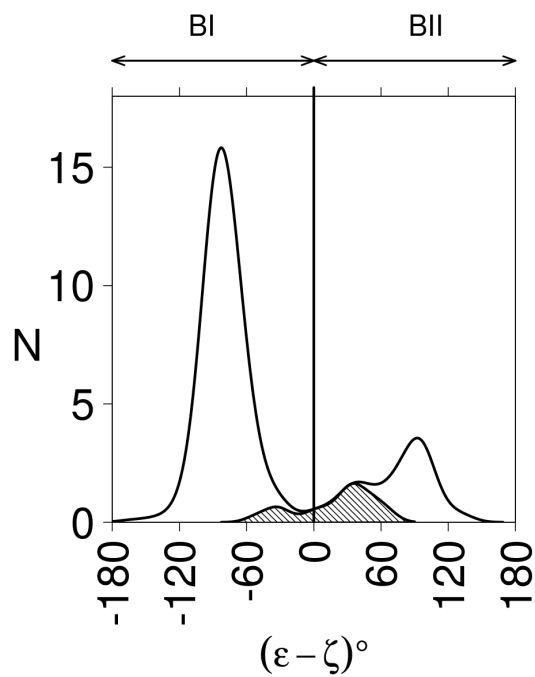
Cumulative BII percentages (BII%) of the dinucleotide steps were extracted from trajectories carried out with either Parmbsc0_{εζOLI} (left, blue) or CHARMM36 (right, red) force fields. BII% were calculated for increasing length of MDs. The BII% were clearly not stabilized during the first 50ns, which are not considered in this analysis. To help identification of the facing steps, sequence in Strand 1 is shown from 5' to 3', while sequence in Strand 2 is shown from 3' to 5'.

Oligo 4



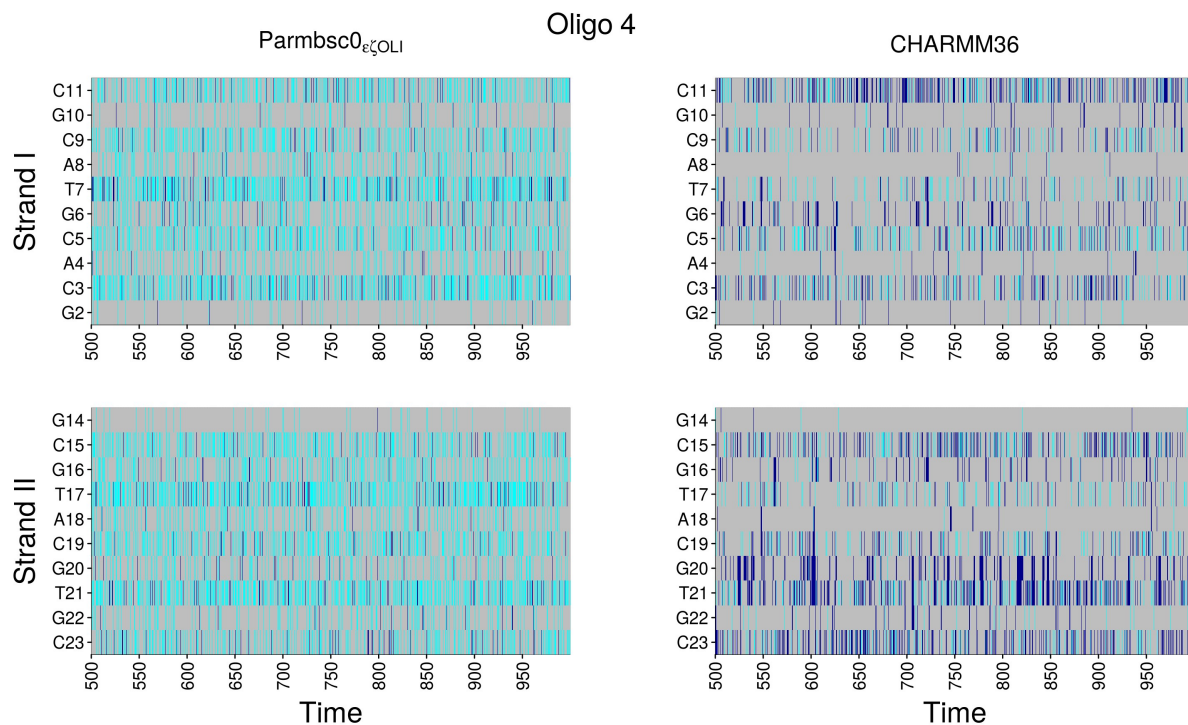
Cumulative BII percentages (BII%) of the dinucleotide steps were extracted from trajectories carried out with either Parmbsc0 $\epsilon\zeta$ OLI (left, blue) or CHARMM36 (right, red) force fields. BII% were calculated for increasing length of MDs. The BII% were clearly not stabilized during the first 50ns, which are not considered in this analysis. To help identification of the facing steps, sequence in Strand 1 is shown from 5' to 3', while sequence in Strand 2 is shown from 3' to 5'.

S3 Fig. (ϵ - ζ) distribution in X-ray structures.



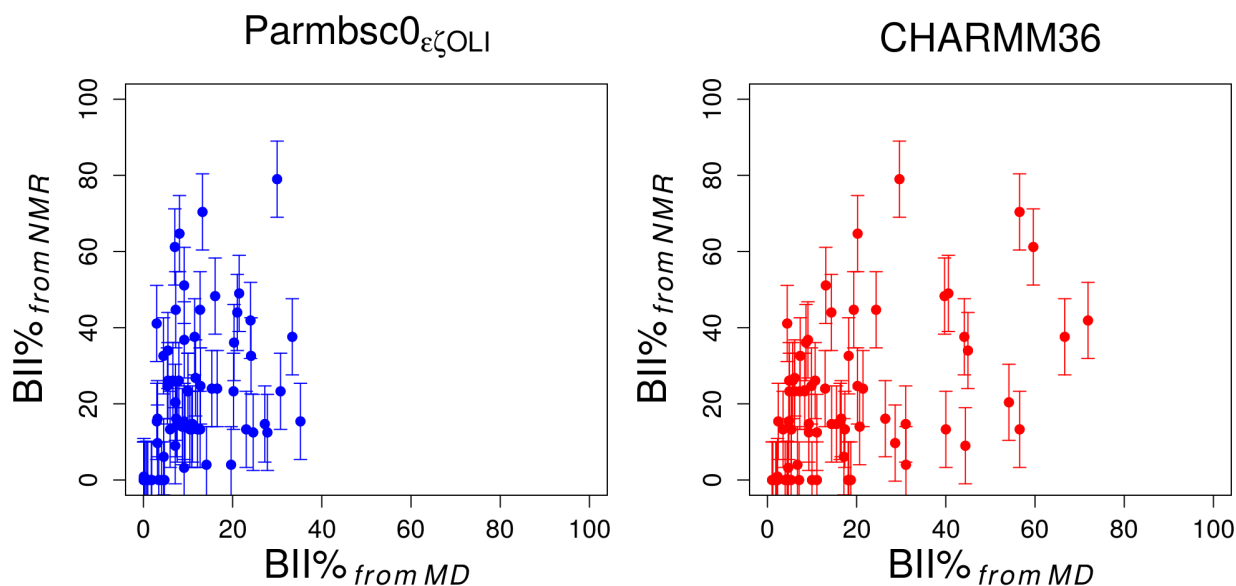
The frequencies (N) of (ϵ - ζ) values ($^{\circ}$) were extracted from 431 steps of high resolution X-ray structures of free DNA. The hachured region corresponds to the distribution of ϵ/ζ :trans/trans.

S4 Fig. Influence of the force-field on the sugar puckers during the MD simulation of Oligo 4.



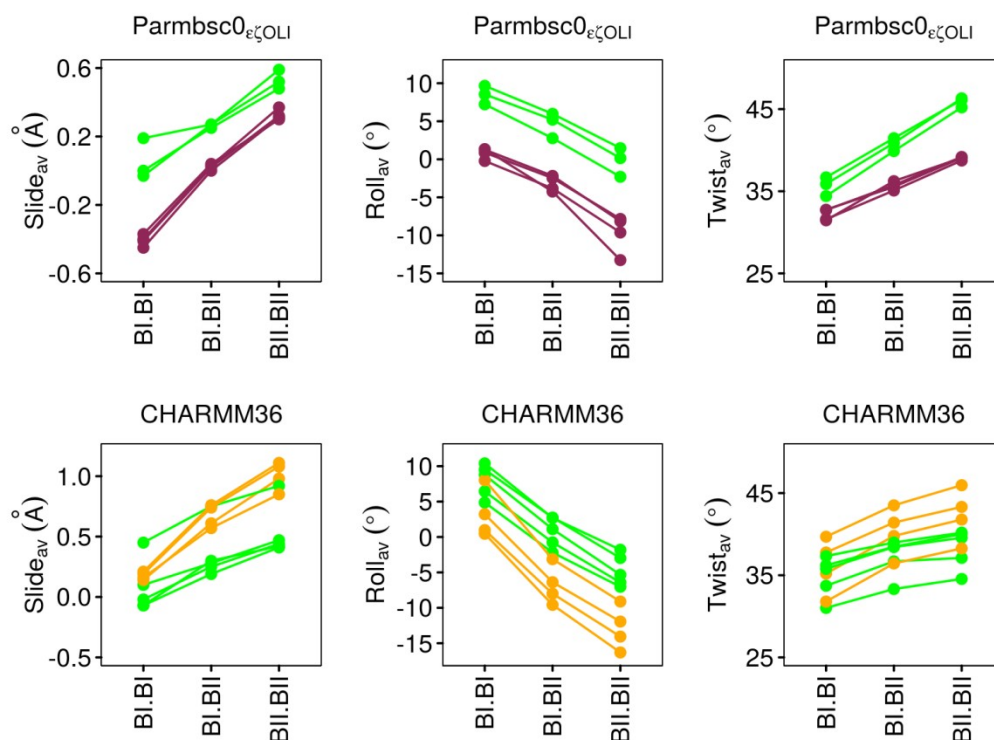
In MDs, the sugars interchange between three puckers, south (grey), east (cyan) and north (blue). The evolution of these conformations versus time (ns) is illustrated for the sugars along the two strands of Oligo 4 in the second part of μ s MDs carried out with Parmbsc0_{eOLi} (left panels) or CHARMM36 (right panels). The sugar puckers are influenced by the force-field (see also Fig 2).

S5 Fig. Comparison between simulated and experimental BII percentages.



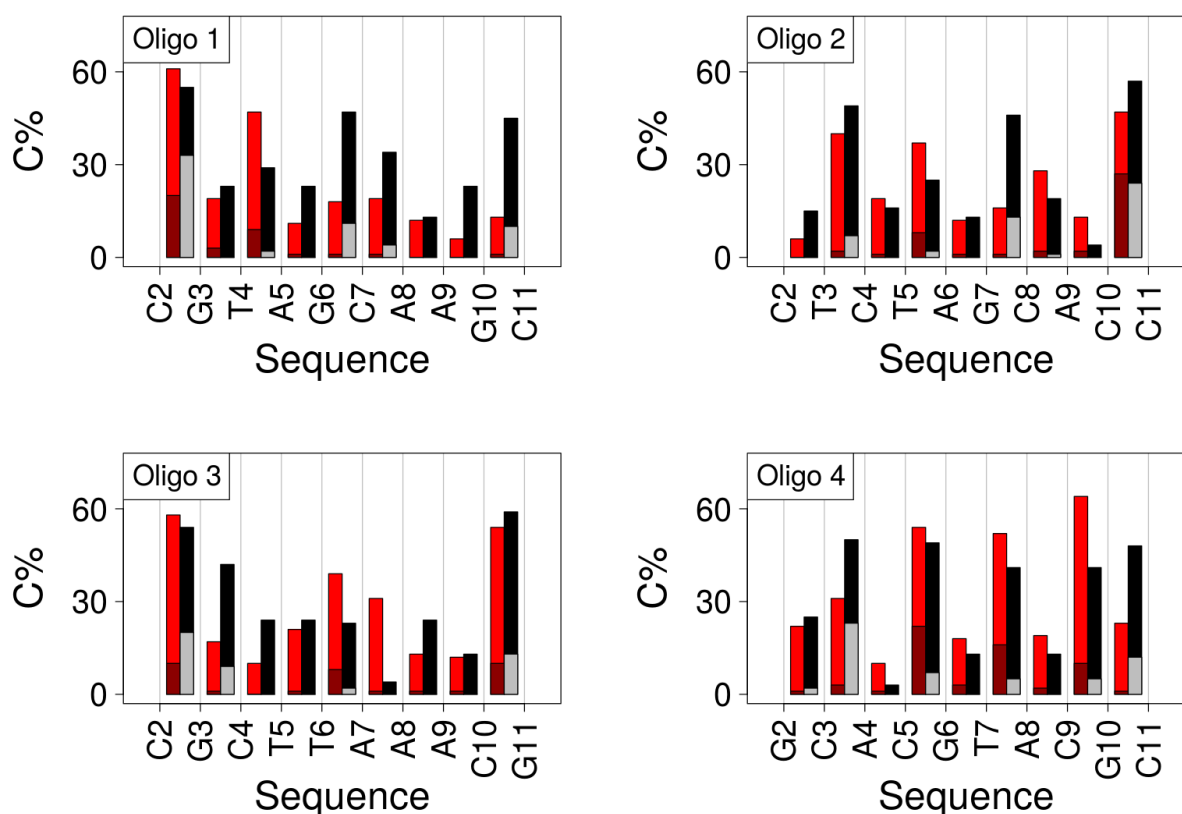
BII percentages were extracted from MDs ($BII\%_{from\ MD}$) with Parmbsc0_{εζOLI} (left panels, blue points) or CHARMM36 (right panels, red points) and compared with the corresponding values inferred from δP s ($BII\%_{from\ NMR}$). The vertical bars represent the error on $BII\%_{from\ NMR}$. The correlation coefficients are given in Table 2 (text section “Sequence-dependent BII propensities from simulations versus NMR.”).

S6 Fig. Slide, Roll and Twist values associated to conformational combinations of facing phosphate linkages in representative BII-rich steps, generated by Parmbsc0_{εζOLI} and CHARMM36.



The values of Slide, Roll and Twist were calculated for representative BII-rich steps, CpG•CpG of Oligos 1, 3 and 4 (green), GpC•GpC of Oligos 1, 2, 3 and 4 (violet) and TpA•TpA of Oligos 1, 3 and 4 (orange) according to the BI·BI, BI·BII|BII·BI and BII·BII combinations of their facing phosphate groups. The data were extracted from C-MDs (top panels) and P-MDs (bottom panels), and averaged along the MD for each backbone state. For clarity, the standard deviations were omitted.

S7 Fig. Comparison between the conformational combinations of facing phosphate linkages extracted from C-MDs or based on NMR data.



The percentages of BI•BII|BII•BI and BII•BII combinations (C%) of facing phosphates are plotted along the four dodecamer sequences. These percentages were extracted from C-MDs (red and dark red bars, stacked, for BI•BII|BII•BI and BII•BII, respectively). Here, they were compared with the percentages calculated in the regime of independence of the conformational states of facing phosphates with equations (2) and (3) (text section “New insights from the simulations: independence of the states of facing phosphate groups.”) and the experimental data (black and grey bars, stacked, for BI•BII|BII•BI and BII•BII, respectively).

S1 Table. BII percentages from NMR and simulations with Parmbsc0_{ε_{COLI}} and CHARMM36.

	Strand 1	BII% _{from} NMR	BII% _{from} P-MD	BII% _{from} C-MD	Strand 2	BII% _{from} NMR	BII% _{from} P-MD	BII% _{from} C-MD
Oligo 1	C2pG3	79	30	30	C22pG23	42	23	75
	G3pT4	0	4	19	A21pC22	23	17	6
	T4pA5	9	7	46	T20pA21	25	14	20
	A5pG6	23	31	7	C19pT20	0	3	6
	G6pC7	26	8	6	G18pC19	44	21	14
	C7pA8	16	7	16	T17pG18	26	10	5
	A8pA9	13	24	10	T16pT17	0	1	1
	A9pG10	23	20	5	C15pT16	0	1	1
	G10pC11	41	3	4	G14pC15	25	11	10
Oligo 2	C2pT3	0	0	6	A22pG23	15	14	2
	T3pC4	15	3	5	G21pA22	48	20	40
	C4pT5	0	1	5	A20pG21	15	38	17
	T5pA6	15	11	33	T19pA20	14	10	20
	A6pG7	13	28	9	C18pT19	0	0	4
	G7pC8	37	9	10	G17pC18	36	21	8
	C8pA9	15	9	14	T16pG17	6	10	13
	A9pC10	4	14	6	G15pT16	0	1	8
	C10pC11	34	6	43	G14pG15	70	43	57
Oligo 3	C2pG3	61	7	59	C22pG23	33	3	17
	G3pC4	33	24	7	G21pC22	26	5	11
	C4pT5	1	0	2	A20pG21	23	11	8
	T5pT6	0	0	2	A19pA20	24	21	20
	T6pA7	16	3	25	T18pA19	10	3	29
	A7pA8	4	20	31	T17pT18	0	0	3
	A8pA9	24	16	13	T16pT17	0	0	2
	A9pC10	13	6	4	G15pT16	0	4	11
	C10pG11	20	7	53	C14pG15	65	10	20
Oligo 4	G2pC3	15	8	17	G22pC23	15	24	9
	C3pA4	45	13	24	T21pG22	51	11	15
	A4pC5	3	9	6	G20pT21	0	2	8
	C5pG6	49	21	40	C19pG20	13	30	55
	G6pT7	0	2	19	A18pC19	13	12	6
	T7pA8	13	12	38	T17pA18	38	14	45
	A8pC9	13	10	5	G16pT17	0	5	19
	C9pG10	38	33	67	C15pG16	13	15	18
	G10pC11	27	12	7	G14pC15	45	7	18

The BII percentages of the individual phosphates were inferred from δP s (BII%_{from} NMR, with a tolerance of $\pm 10\%$) or extracted from the simulations carried out with Parmbsc0_{ε_{COLI}} (BII%_{from} P-MD) or CHARMM36 (BII%_{from} C-MD).

S2 Table. Populations of the three conformational combinations of facing phosphate groups based on experimental data.

		BI•BI%	BI•BII BI•BII%	BII•BII%
Oligo 1	C2pG3•C22pG23	12	55	33
	G3pT4•A21pC22	77	23	0
	T4pA5•T20pA21	69	29	2
	A5pG6•C19pT20	77	23	0
	G6pC7•G18pC19	41	47	11
	C7pA8•T17pG18	62	34	4
	A8pA9•T16pT17	88	13	0
	A9pG10•C15pT16	77	23	0
	G10pC11•G14pC15	44	45	10
Oligo 2	C2pT3•A22pG23	85	15	0
	T3pC4•G21pA22	44	49	7
	C4pT5•A20pG21	84	16	0
	T5pA6•T19pA20	73	25	2
	A6pG7•C18pT19	88	13	0
	G7pC8•G17pC18	40	46	13
	C8pA9•T16pG17	80	19	1
	A9pC10•G15pT16	96	4	0
	C10pC11•G14pG15	20	57	24
Oligo 3	C2pG3•C22pG23	26	54	20
	G3pC4•G21pC22	49	42	9
	C4pT5•A20pG21	76	24	0
	T5pT6•A19pA20	76	24	0
	T6pA7•T18pA19	76	23	2
	A7pA8•T17pT18	96	4	0
	A8pA9•T16pT17	76	24	0
	A9pC10•G15pT16	87	13	0
	C10pG11•C14pG15	28	59	13
Oligo 4	G2pC3•G22pC23	73	25	2
	C3pA4•T21pG22	27	50	23
	A4pC5•G20pT21	97	3	0
	C5pG6•C19pG20	44	49	7
	G6pT7•A18pC19	87	13	0
	T7pA8•T17pA18	54	41	5
	A8pC9•G16pT17	87	13	0
	C9pG10•C15pG16	54	41	5

The percentages of the BII•BII, BI•BII|BII•BI and BII•BII combinations were calculated for the four dodecamers from the BII percentages of individual phosphates, BII%_{from NMR}, (see S1 Table) using the equations (2) and (3) given in the text section “New insights from the simulations: independence of the states of facing phosphate groups.”.

S3 Table. Values of inter base pair parameters according to the conformational combinations of facing phosphate groups.

	Shift		Tilt		Rise	
	P-MDs	C-MDs	P-MDs	C-MDs	P-MDs	C-MDs
BI•BI	0 (0.1)	0 (0.5)	-0.2 (1.5)	0 (5)	3.3 (0.1)	3.4 (0.4)
BI•BII BII•BI	0 (0.5)	0 (0.6)	-0.4 (2.4)	-0.1 (5.5)	3.4 (0.1)	3.5 (0.4)
BII•BII	0 (0.2)	0 (0.5)	-0.5 (2.6)	-0.1 (5.7)	3.5 (0.2)	3.6 (0.5)

The characteristic values of Shift (\AA), Tilt ($^\circ$) and Rise (\AA) with standard deviations in brackets are given for the three combinations of facing phosphate groups, regardless the dinucleotidic sequence. The inter base pair parameter values were extracted from the analysis of the simulations with Parmbsc0_{εCOLI} (P-MDs) and CHARMM36 (C-MDs).

CHAPITRE 3

Mécanique de l'ADN-B en solution

1 Introduction

Dans ce chapitre, notre but est de montrer comment on peut mettre en évidence la mécanique de l'ADN en solution. Pour ce faire, nous allons utiliser des couplages dipolaires résiduels que j'ai mesurés sur nos quatre dodécamères. Avant de résumer notre démarche et les résultats, je ferais une brève introduction sur ces quantités mesurables en RMN.

En présence d'un champ magnétique, chaque noyau possédant un moment magnétique de spin non nul se comporte comme un dipôle magnétique. L'interaction entre les dipôles magnétiques I et S appelée aussi couplage dipolaire a une amplitude D_{\max} donnée par l'équation (30)

$$D_{\max} = -\frac{h\gamma_I\gamma_S\mu_0}{8\pi^3 r_{IS}^3} \quad (30)$$

où h est la constante de Planck, γ_I et γ_S sont les rapports gyromagnétiques des spins I et S respectivement, r_{IS} est la distance entre les deux noyaux et μ_0 est la constante de perméabilité magnétique du vide. L'amplitude de l'interaction dipolaire (D_{\max}) d'après l'équation (30) est une constante pour chaque couple de noyaux. Dans les spectres RMN, les couplages dipolaires devraient apparaître sous forme de résonances séparées par la quantité ($\Delta\nu$), analogue à l'effet des couplages scalaires. Toutefois, cette séparation des résonances ne se manifeste en aucun cas dans les spectres des molécules en solution isotrope. En effet, $\Delta\nu$ est directement relié à l'angle formé par le vecteur internucléaire IS et le champ magnétique statique du spectromètre B_0 (Figure 27), selon

l'équation (Bax, Kontaxis, et Tjandra 2001) : $\Delta\nu = D_{\max} \frac{1}{2} \langle 3 \cos^2 \theta - 1 \rangle \quad (31)$

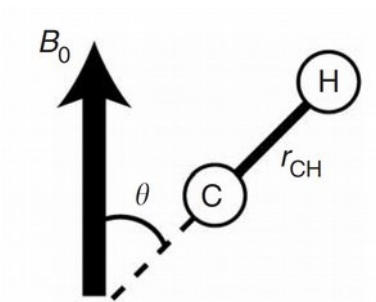


Figure 27: Représentation schématique de l'orientation du vecteur internucléaire reliant deux dipôles magnétiques carbone (C) et proton (H), de distance r_{CH} et formant un angle θ avec champ magnétique statique B_0 .

Le terme $\langle 3 \cos^2 \theta - 1 \rangle$ s'annule en milieu isotrope, puisque il s'agit d'une moyenne de toutes les orientations possibles du vecteur IS dans l'espace et dans le temps. La mesure de la l'interaction dipolaire n'est possible que dans un milieu cristal liquide, où les macromolécules ont des orientations préférentielles. Dans ce cas le terme $\langle 3 \cos^2 \theta - 1 \rangle$ n'est pas nul : on parle alors de couplage résiduel dipolaire ou RDC (*Residual Dipolar Coupling*).

Dans les macromolécules, les RDC peuvent être exprimés selon les trois composantes principales du tenseur d'alignement de la molécule (A_{xx} , A_{yy} et A_{zz}) et les angles que forme le vecteur internucléaire IS avec les composantes z (ζ) et x (ϕ) (équation 32 et Figure 28)

(Bax, Kontaxis, et Tjandra 2001).

$$\Delta\nu = D_{max} [A_a (3 \cos^2 \zeta - 1) + A_r (\sin^2 \zeta \cos 2 \Phi)] \quad (32) \quad \text{où} \quad A_{aa} = \frac{1}{2} A_{zz} \quad A_r = \frac{1}{3} (A_{xx} - A_{yy}) \quad (33)$$

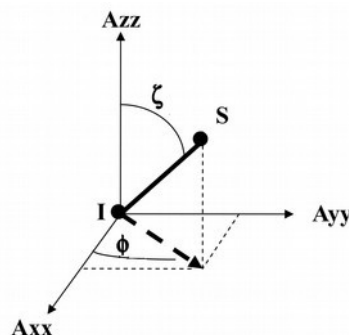


Figure 28 : Vecteur internucléaire IS représenté dans son tenseur d'alignement A dont les composantes principales sont A_{xx} , A_{yy} et A_{zz} (selon les axes x, y et z) et les angles ζ et ϕ formés entre le vecteur IS et les composantes z et x.

Les mesures de RDC sont souvent effectuées sur les couples de noyaux NH et CH reliés par une

unique liaison chimique. Sachant que les distances r_{is} dans ces cas sont fixes et connues, les RDC dépendent uniquement de l'orientation des vecteurs internucléaires par rapport à l'axe moléculaire. Dans le cas de dodécamères d'ADN en double hélice B, le terme A_r est très faible (Wu et al. 2006) donc les RDCs sont régies essentiellement par l'angle ζ .

Dans les publications antérieures concernant les ADN, les RDC ont été utilisés comme contraintes, en plus des distances, lors des affinements de structures par RMN (Tjandra et al. 2000, MacDonald et al. 2001, Mauffret, Tevanian, et Fermandjian 2002). Dans le travail présenté ici, nous avons pour premier but de sonder les interdépendances qui peuvent exister entre les observables RMN au niveau des dinucléotides. Nous avons donc considéré les différences entre deux RDC consécutifs (ΔRDC). Ces ΔRDC constituent ainsi un jeu de données cohérent avec les distances internucléotidiques et les déplacements chimiques du phosphore.

Nous avons ainsi pu mettre en évidence des corrélations entre les δP et les ΔRDC impliquant les vecteurs C6/8-H6/8, C3'-H3' et C4'-H4' d'une part et les δP et les distances internucléotidiques impliquant les protons H6/8, H2' and H2'' d'autre part. Pour pouvoir interpréter structuralement ces corrélations, nous avons calculé leurs équivalents dans les structures cristallographiques puisque ces dernières offrent un accès direct aux paramètres hélicoïdaux. Les distances internucléotidiques et les ΔRDC calculés sur les structures cristallographiques se sont montrés sensibles aux paramètres hélicoïdaux de *Twist*, *Roll* et *Slide*. Le parallèle fait entre les données RMN et celles recalculées sur les structures cristallographiques nous ont permis de comprendre le lien entre états conformationnel BI-BII et les paramètres *Twist*, *Roll* et *Slide* en solution. Les analyses des structures cristallographiques ont révélé le lien étroit qui existe entre les paramètres hélicoïdaux et les états conformationnels BI-BII (Djuranovic et Hartmann 2003) (Djuranovic et Hartmann 2004), cependant la validité de ces relations en solution n'était pas confirmée expérimentalement. Les résultats présentés dans ce chapitre constituent la première preuve expérimentale de l'existence d'une mécanique équivalente à celle trouvée dans les structures cristallographiques, qui montrent que les mouvements des groupements phosphate et ceux bases aromatiques sont hautement coordonnés en solution.

Un volet de nos travaux est consacré aux comportements des paires de bases terminales. Lors des dynamiques moléculaires présentées dans le chapitre précédent, nous avons vu que les bases terminales s'ouvraient sans se refermer et engageaient des interactions variées avec les bases et les

phosphates des dinucléotides adjacents. Ces multiples événements, d'une durée variable, affectent considérablement la dynamique de la région proche des bases terminales.

Nous avons donc revisité l'ensemble de nos données RMN pour sonder les comportements des bases terminales des quatre oligomères. Bien que les résultats tels qu'ils sont présentés dans ce chapitre ne visent pas directement à comparer les dynamiques moléculaires et les données RMN, j'ai pris un exemple (Tableau 1) de la paire de base G12:C13 de l'oligomère 4 pour montrer que le comportement de cette paire de bases dans les simulations n'a aucun fondement expérimental et qu'elle est un artefact du champ de force.

		Distance (Å)		Angles	
		H2' _{C11} -H8 _{G12}	H2'' _{C11} -H8 _{G12}	α	χ
RMN	C ₁₁ pG ₁₂	3.7	3.2	<i>g</i>-	anti
DM	C ₁₁ pG ₁₂	8.3	7.4	<i>t</i>	syn

Tableau 4 : Comparaison des distances inter-nucléotidiques H2'_{C11}-H8_{G12} et H2''_{C11}-H8_{G12}, des conformations de l'angle de brin α et de l'angle glycosidique χ du G12 issus des données RMN et les moyennes extraites de la dynamique moléculaire réalisée avec Parmbsc0_{εζOLI}. Le pas considéré ici est l'un des pas terminal de l'oligomère 4.

Les distances H2'_{C11}-H8_{G12} et H2''_{C11}-H8_{G12} moyennées sur la trajectoire de la dynamique moléculaire sont si grandes qu'elles ne pourraient jamais apparaître en RMN. De même, pour l'angle α du brin et l'angle χ , les données expérimentales ne sont pas compatibles avec les données de la dynamique moléculaire.

Les données RMN des résidus terminaux dont nous disposons des quatre oligomères montrent que :

- i- tous les NOE H1'-H6/8 intra-résidu vont dans le sens d'angles glycosidiques χ en conformation anti ;
- ii- tous couplages scalaire des H4'-P de tous les résidus terminaux en 3' de chaque brin sont visibles à 20, 30 et 40° C, témoignant d'angles de brin α , β et γ canonique, autrement ces couplages scalaires ne seraient pas visibles ;
- iii- tous les déplacements chimiques du phosphore à 20, 30 et 40° C des résidus terminaux sont comparables à ceux des résidus internes ;
- iv- les distances internucléotidiques et les RDC impliquant les résidus terminaux sont dans les

gammes de distances et des RDC des résidus internes.

Bien que les protons des groupements imino des thymines et des guanines lorsqu'elles sont en position terminales ne soient pas visibles à 20°C, témoignant que ces bases ne sont pas engagées dans des liaisons hydrogène, toutes les autres données RMN montrent que les mouvements associés à ces ouvertures gardent en moyenne des caractéristiques proches de l'ADN-B canonique.

2 Article 3

The intrinsic mechanics of B-DNA in solution characterized by NMR

**Akli Ben Imeddourene , Xiaoqian Xu , Loussiné Zargarian, Christophe Oguey,
Nicolas Foloppe, Oliver Mauffret and Brigitte Hartmann**

Soumis à *Nucleic Acids Research*

The intrinsic mechanics of B-DNA in solution characterized by NMR

Akli Ben Imeddourene^{1,2,§}, Xiaoqian Xu^{1,3,§}, Loussiné Zargarian¹, Christophe Oguey⁴, Nicolas Foloppe⁵, Oliver Mauffret^{1,*} and Brigitte Hartmann^{1,*}

¹ Laboratoire de Biologie et Pharmacologie Appliquée, ENS Cachan, CNRS, Université Paris-Saclay, 61 avenue du Président Wilson, 94235 Cachan cedex, France;

² Université Pierre et Marie Curie, Paris, France;

³ Department of Life Sciences, East China Normal University, 200062 Shanghai, People's Republic of China;

⁴ [Laboratoire de Physique Théorique et Modélisation](#), CNRS, Université de Cergy-Pontoise, Cergy-Pontoise, France;

⁵ 51 Natal Road, Cambridge CB1 3NY, UK.

* To whom correspondence should be addressed. Tel: +33 147 407 421 or +33 147 407 733; Fax: +33 147 407 671; Emails: bhartman@ens-cachan.fr, olivier.mauffret@lbpa.ens-cachan.fr

Keywords: DNA structure and dynamics; NMR of nucleic acids; RDC; BI/BII equilibrium.

ABSTRACT

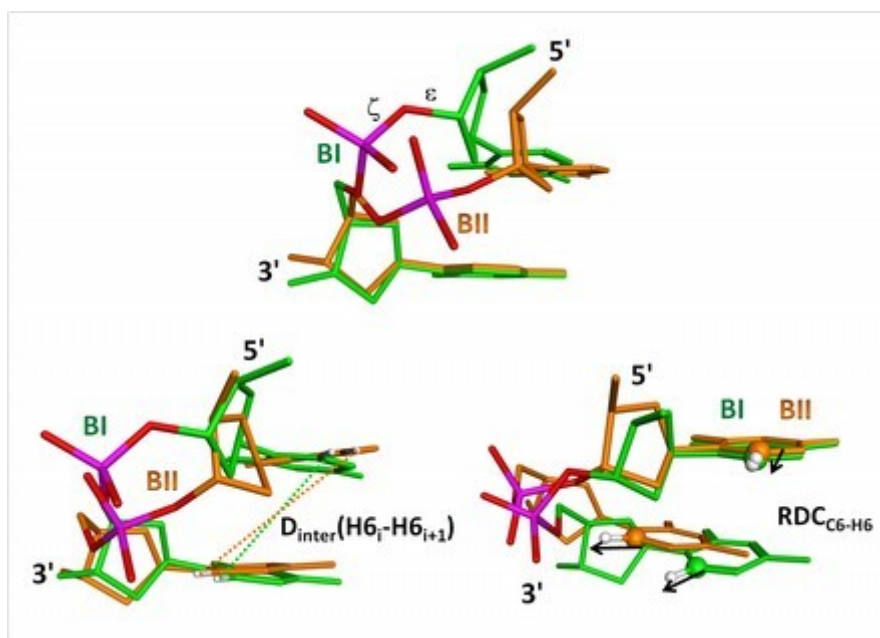
An experimental characterization in solution of the structural couplings in free DNA has been elusive, since they are subtle and have proved difficult to tackle. Here, the direct exploitation of the NMR measurements collected on four B-DNA dodecamers provides a consistent series of correlations that offers a novel way to characterize the intrinsic mechanics of free DNA in solution. The difference between two successive Residual Dipolar Couplings (Δ RDCs) involving C6/8-H6/8, C3'-H3' and C4'-H4' vectors, and internucleotide distances involving H6/8, H2' and H2'' protons are correlated to ³¹P chemical shifts (δ P), which reflect the populations of the BI and BII backbone states. Calculations of NMR quantities on very high resolution X-ray and geometrically controlled models of DNA structures show for the first time that Δ RDCs depend mostly on roll while internucleotide distances are mainly sensitive to twist or slide. These relations uncover extensive and consistent evidence characterizing the intrinsic B-DNA mechanics, and illustrate how δ P measurements can qualitatively inform on key inter base pair parameters, in addition to probe the BI \leftrightarrow BII backbone equilibrium. The inspection of the 5' and 3' ends of the dodecamers also provides new information on the fraying events, otherwise neglected.

[§]The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors

INTRODUCTION

Characterizing the relationships between the different structural parameters describing the B-DNA is crucial to understand DNA malleability and interactions with other molecules. Indeed, the DNA behavior and intrinsic mechanical properties is a fundamental, but still poorly understood, aspect of DNA-protein complex formation, especially in DNA indirect readout. An earlier analysis of a large dataset of X-ray structures showed that in both free and bound DNA the slide, roll and twist are correlated to various degrees, depending on the dinucleotidic sequence (1). Several other studies, based on high resolution X-ray structures, revealed intimate structural couplings between the phosphate group states, sugar conformations, inter base pair parameters, base pair displacement (X-disp) and groove dimensions (2-9). The backbone conformations in free B-DNA consist of two states, BI and BII, originally defined by the conformations of the torsion angles ϵ and ζ , *trans/g-* in BI (ϵ - $\zeta \sim -90^\circ$) and *g-/trans* in BII (ϵ - $\zeta \sim +90^\circ$) (10) (Figure 1). Analysis of these states enabled to define two structural profiles at the dinucleotide level: pure BI linkages correspond to negative slide, null or positive roll and low or moderate twist (BI profile), while pure BII linkages are associated with positive slide, negative roll and high twist (BII profile) (2,3). The minor and major groove dimensions are also intrinsically coupled to the BII density (4,11). Several molecular modeling studies have supported this view of intrinsic DNA mechanics (3,12-15).

Figure 1: Structural basis for the relation between BI and BII states and two types of NMR observables.



In each panel, the BI and BII states are compared by superposition of a CpC dinucleotide in these two backbone conformations (X-ray structure, PDB ID: 3GGB). CpC in BI (sugars and bases in green) is characterized by a roll of $+8.5^\circ$ and a twist of 28° , while CpC in BII (sugars and bases in orange) has a roll of -2° and a twist of 45° . Top: comparison of the BI and BII phosphate linkage conformations resulting from changes in the ϵ and ζ torsions. Bottom left: dependence of the

internucleotide distances on the phosphate state, illustrated with a H6-H6 internucleotide distances which is longer in a BII (orange dashed lines) than in a BI step (green dashed lines). Bottom right: influence of the backbone conformation on the relative orientation between internucleotides RDCs, illustrated with C6-H6 vectors (RDC_{C6-H6}) indicated by black arrows.

The knowledge about DNA intrinsic structural couplings would advance decisively if it could rely on experimental data in solution, in addition to high quality crystals. That could take advantage of residual dipolar couplings (RDC) obtained from Nuclear Magnetic Resonance spectroscopy (NMR), since their introduction in refinements may substantially ameliorate the accuracy of DNA structures (16-19). RDCs are collected in partially oriented media and derive from an incomplete averaging of anisotropic magnetic interactions between proximal active spins of nuclei, typically carbons and hydrogen in C-H bonds in DNA. They inform about the angle between specific C-H vectors and the alignment tensor, which is roughly parallel to the external applied magnetic field B_0 when Pf1 phages are used to orientate the DNA (20,21). Such information complements the other NMR data, essentially internucleotide distances from Nuclear Overhauser Effect (NOE) connectivities, useful to constrain the relative position of two successive nucleosides, and scalar couplings to investigate the behavior of sugars and phosphodiester linkages. Thus, the recent DNA structural models obtained from NMR are expected to be better than more loosely restrained models for exploring the B-DNA mechanics.

However, the DNA structures obtained from RDC and NOE restraints (16,18,22-28) remain scarce comparing to protein structures and suffer from several limitations. The sequences are dominated by A-tracts (17,23,24,28) and the Drew-Dickerson dodecamer (25-27). Such redundancy has revealed inconsistencies between the refined structures of oligomers of similar or identical sequences. For instance, TpA in T_3A_3 or T_4A_4 had either a negative (17) or a marked positive roll (28); the structures of the Drew-Dickerson dodecamer show substantial differences when refined from the same NMR dataset but with different NMR-based backbone restraints (25-27). More generally, even if the restraints are expected to limit the impact of refinement protocols, the different force fields and representations of solvent (implicit or explicit) raise the issue of the reproducibility and reliability of those NMR DNA structures. The very short durations of the restrained molecular dynamics, 1ns or less, and the recurrent use of artificial restraints on the backbone angles (17,23,26-28) could also introduce biases on the final DNA representation. In sum, the lack of sequence variety combined with limitations of the refinement protocols have not yet delivered a dataset of structures with consistent information on the intrinsic mechanics of free B-DNA in solution.

Here, we approach this intrinsic mechanics by exploiting directly the NMR measurements instead of the derived refined structural models. In addition to RDCs and distances, ^{31}P chemical shift (δP) measurements were also taken into account. The structural significance of δP has been deciphered earlier, mainly from correlations between δP and coupling constants (29,30), and more recently from quantum chemistry calculations (31,32). Thus, the dispersion of δP in B-DNA is attributed to the variable propensity of phosphate groups to undergo the BI \leftrightarrow BII conformational exchange. In solution, high-field shifted δP indicate BI-rich linkages while down-field shifted δP are

typical of BII-rich linkages (29,30,32). In practice, the translation of δP in terms of BI/BII ratios is now possible using two different strategies (33,34).

The δP s can be used to probe the backbone behavior and its strong sequence dependence (11,25,33-35). Taking advantage of the accuracy of δP measurements, it has emerged that specific δP values, and hence specific BI/BII ratios, characterize the 16 unique dinucleotides. These specific δP values led to conceive an experimental scale in which the 10 complementary dinucleotides are characterized by individual scores that represent their BII propensities (36). This scale is called TRX in reference to the relationship between BI and BII states, Twist, Roll and base pair displacement (X-disp), with reference to the BI and BII profiles from X-ray data, as explained above. This view is supported in solution by consistent NMR data collected on a 14bp DNA, in which internucleotide distances were found to be correlated to δP s (33).

The aim of the present study is to further develop this approach, using a large variety of NMR data, comprising δP s, distances and RDCs, collected on four non-palindromic dodecamers that offer a substantial set of dinucleotide sequences. These oligomers, once juxtaposed, correspond to a fragment of the “sequence 601”, known for its strong ability to form nucleosomes (37). Recently, the fresh δP s from the four dodecamers were used as a test set, which validated the TRX scale, thereby demonstrating that the BII propensities are predictable (11).

To broaden and deepen the previous analyses, here we uncover and interpret relationships between δP s on one hand, and ΔRDC s and internucleotide distances on the other hand. ΔRDC s are the differences between two RDCs of successive bases on the same strand. ΔRDC s have the advantage to be directly related to the angles between two dipolar vectors, avoiding the necessity to calculate the alignment tensor. Indeed, their equivalents are used in protein refinement protocols (38,39). We show that when the ΔRDC s are considered with δP s and internucleotide distances, this combination provides a consistent data package which offers a novel way to characterize the behavior of the dinucleotides in free DNA in solution. We then determined which helical parameters of DNA influence the ΔRDC s and internucleotide distances, using ultrahigh resolution X-ray structures as experimental models. In the last part of our study, the 3'- and 5'-terminal dinucleotides were scrutinized to determine if the fraying events, still experimentally poorly characterized, perturb the double helix as severely as suggested by recent MD simulations (12,40,41).

Overall, the results offer new, extensive characterization of the B-DNA intrinsic mechanics in solution. They imply that the local helicoidal parameters can be estimated from a readout of δP s, which could be advantageous considering that the δP s can be accessed with relative ease.

MATERIALS AND METHODS

DNA sequences

Four oligodeoxyribonucleotides of 12 base pairs (bp) (sequences in Table 1) were studied by NMR. These sequences, placed end to end after discarding the terminal base pairs, allow to recompose a continuous 39 bp segment corresponding to the 5' part of the non-palindromic sequence 601, selected from SELEX experiments for its very high-affinity for association with the histone octamers (37).

Sample preparation

Oligomers were synthesized by Eurogentec Inc. (Belgium). The sample was dissolved in an aqueous sodium-phosphate buffer corresponding to an ionic strength of 0.1 (mol/L) with 0.1 mM EDTA, at pH 6.5. The duplexes were prepared by mixing the two complementary strands in a 1:1 ratio in 450 μ L H₂O and 50 μ L D₂O for studies of exchangeable protons. For studies of non exchangeable protons, the duplexes were lyophilized three times and dissolved in 500 μ L of 99.9% D₂O. The final concentration of double strand oligomers is 1-1.5 mM.

NMR spectroscopy

³¹P chemical shifts and distances

All NMR spectra were recorded on a Bruker Avance spectrometer operating at a proton frequency of 500 MHz and at a phosphorus frequency of 202 MHz with a 5-mm gradient indirect probe. All spectra were processed with NMRpipe (42) or Topspin softwares (Bruker) and analyzed with Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

One-dimensional ¹H spectra collected from 5 to 60 °C with a 5 °C step enabled to check the stability of the duplexes over this range of temperature. All the imino protons were clearly observable until 40 °C, except those of the terminal base pairs.

¹H NMR studies were performed at 10, 20 and 30 °C. 2D NOESY spectra were recorded using mixing times of 100, 200, 300 ms for exchangeable protons and 80, 100, 200, 300 and 400 ms for non exchangeable protons. MLEV-17 TOCSY experiments were run using mixing time of 120 ms. The 2D NOESY and MLEV-17 TOCSY were recorded under the following experimental conditions: 2048 data points and 512 t₁ increments with spectral widths of 10,000 Hz for exchangeable protons and 5,000 Hz for non exchangeable protons. The water signal was suppressed with a WATERGATE sequence (43). NOE distances were extracted with particular care from cross-peaks after visual inspection of the build-up rates and using the distance extrapolation method to correct the spin-diffusion effects (44). However, some large distances (typically > 5 Å) were only observable in NOESY spectra recorded with a mixing time of 300ms and were thus less accurately measured than shorter distances. Distances were normalized using the cytosine H5–H6 proton pairs ($r = 2.5$ Å) apart from those involving the CH₃ group, for which the reference was the average H6–CH₃ distance ($r = 2.9$ Å). The experimental error was estimated to 10 percent of the measured distances.

¹H-³¹P HETCOR (45) experiments were run at 20, 30, 40 °C. The spectra width was 2500 Hz in the ¹H dimension and 810 Hz in the ³¹P dimension. Data were recorded with 2048 points in the ¹H dimension and 256 increments in the ³¹P dimension. ³¹P chemical shifts (δ P) were referenced

relative to internal trimethyl phosphate. Uncertainty of δP s measured in solution was estimated to ± 0.02 ppm. Note that in case of correlations between δP and ΔRDC s, the uncertainty on δP s was estimated to ± 0.05 ppm to take account of the slight variations that could be induced by the oriented medium (46).

All the NMR data presented here were collected at 20°C.

Sugar conformations

The sugar conformation was estimated by inspecting the intensities of TOCSY H1'-H4' cross-peaks. These intensities are related to the values of the successive scalar couplings involved in the TOCSY H1'-H4' transfer, $J_{H1'-H2'}$, $J_{H1'-H2''}$, $J_{H2'-H3'}$, $J_{H2''-H3'}$ and $J_{H3'-H4'}$. Among these scalar couplings, $J_{H3'-H4'}$ is strongly sensitive to the sugar conformation (47,48). Its value, quasi null for *South* (C2'-endo) sugars, increases with *North* (C3'-endo) or *East* (O4'-endo) populations and allows the H1'-H4' transfer. In practice, observable TOCSY H1'-H4' cross-peaks thus sign the presence of *North* or *East* conformers. That TOCSY H1'-H3' cross-peaks of all the residues in the four dodecamers are observed while several TOCSY H1'-H4' cross-peaks are missing confirms the role of bottleneck played by $J_{H3'-H4'}$, as postulated before (49).

Conformation of the backbone angles β and γ

The conformation of the backbone angles β and γ was deduced from the observation of $^4J_{H4'-P}$ coupling constants. These couplings appear in the spectra only if β/γ is in the *trans/g+* region (29,47), typical of B-DNA.

Residual Dipolar Coupling constants

With DNA oligomers, low concentrations (few mg/ml) of Pf1 phages are adequate for forming an ordered medium in which the long axis of the phages aligns parallelly to the magnetic field B0 (50,51). These oriented phages promote the DNA alignment, which is considered to be nearly parallel to B0 (19,50).

Oriented solutions were obtained by adding 12 mg/ml Pf1 phage to oligonucleotide solutions. The phage concentration was chosen to obtain a homogeneous liquid crystal phase and to preserve the spectrum quality, *i.e.* well defined peaks. The solution homogeneity and the extent of orientation have been checked with 2H detection of HDO peak (52).

The measurements were performed at ^{13}C natural-abundance conditions. The ^{13}C - 1H direct Residual Dipolar Coupling constants (RDC) were determined using different strategies for anomeric (C1', C3', C4') and aromatic (C8, C6, C2) carbons. For the anomeric carbons, ^{13}C - 1H RDC were determined by comparing the in-phase ^{13}C - 1H splittings in the direct dimension of non-decoupled 2D HSQC spectra acquired in isotropic and oriented solutions (26). For the aromatic carbons improved sensitivity and resolution could be obtained using TROSY experiments (53,54). Accurate RDC measurements have been performed by recording pairs of complementary H α and H β -TROSY experiments on isotropic and oriented solutions.

TROSY and HSQC experiments were performed at 500 MHz at 25°C. The spectral range was 5000 Hz for proton dimension and 5030 Hz for the carbon dimension. 512 experiments of 2048 complex points were recorded with an accumulation of 256 scans per experiment.

A total of 291 RDCs were collected. The experimental error on RDCs linked to recording conditions and data processing is estimated to 2.5 Hz, in accordance with previous evaluations (16). The estimated error on Δ RDC, which is the difference between two RDCs, was 5 Hz.

NMR data

The NMR data are available in the Biological Magnetic Resonance Bank (BMRB), entry 19222.

B-DNA structural models

The dataset of crystal structures included 9 B-DNA decamers (PDB IDs: 3GGI 1EN8 1DG8 1ENE 3GGB 3GGK 1EN9 1EN3 1ZF5) from data that extends beyond 1.0 Å resolution. The stringent resolution cut-off was chosen to calculate RDCs and internucleotide distances that must be as precise as possible. The protons were added with Amber tools (D.A. Case et al, 2006, AMBER 9, University of California, San Francisco), known to position hydrogen reliably on DNA molecules. All the considered X-ray oligomers correspond to palindromic sequences and their structures are symmetric; this was taken into account to avoid redundancy. The backbone, sugar and base pair positional heterogeneities in 3GGB, 3GGI and 3GGK were considered and analyzed. The final X-ray dataset contained 108 dinucleotide steps.

$D_{\text{inter}}(\text{H6}/8_i\text{-H6}/8_{i+1})$, $D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$, $D_{\text{inter}}(\text{H2}''_i\text{-H6}/8_{i+1})$ were directly inferred from the added protons. RDCs involving C6/8-H6/8, C3'-H3', C4'-H4' vectors were calculated on X-ray structures with PALES (55), using a concentration of Pfl phage of 10 mg/ml.

In addition to the X-ray structures, a series of DNA models were generated with 3DNA (56). They are based on canonical B-DNA structures (called “generic” in 3DNA) with C2'-endo sugars, BI phosphates, twist of 36° and rise of 3.4Å. We considered two sequences, (CpG)₂₀•(CpG)₂₀ and (GpG)₂₀•(CpC)₂₀, to get a representative collection of different purine-pyrimidine types (CpG, GpC, GpG and CpC). In these DNAs, the slide, roll and twist of one dinucleotide were successively and independently scanned over the range of values observed in X-ray structures. For instance, twist was constrained from 20 to 50° by interval of 5°, keeping the other helical parameters at their canonical values.

Structural analyses

Analyses of DNA X-ray structures were carried out using Curves5 (57), which calculates the optimal helical axis and a complete set of conformational and helicoidal parameters. To be consistent with the NMR data, inter base-pair parameters were calculated separately for dinucleotides NpN in each DNA strand and not for complementary dinucleotides NpN•NpN which combine the two strands.

RESULTS

Correlations between NMR observables

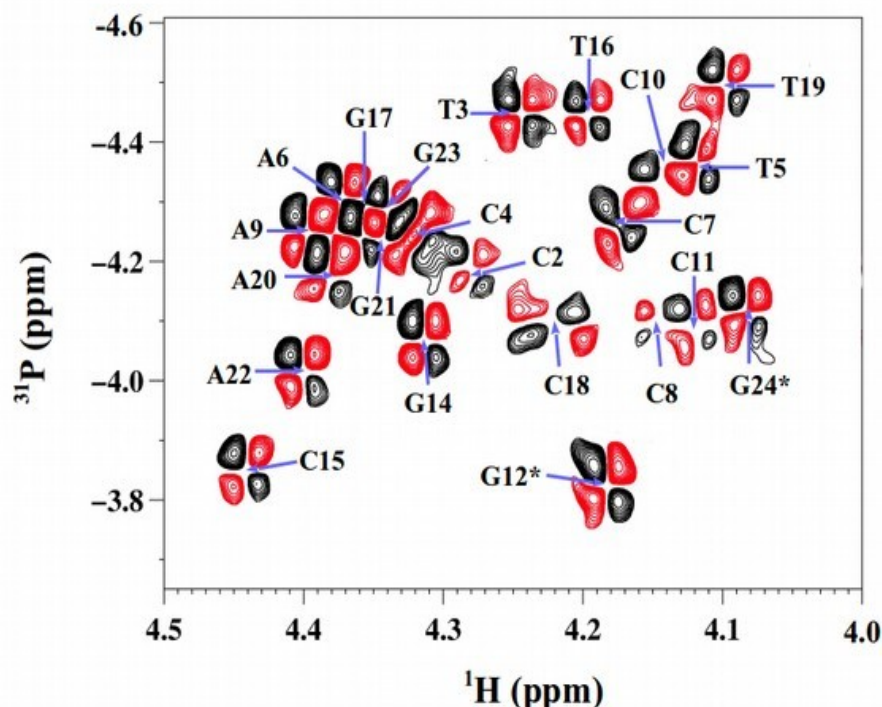
The four dodecamers studied by NMR (Table 1) offer a large representation of YpR, RpY, YpY and RpR (Y:pyrimidine; R: purine) steps. 72 ^{31}P chemical shifts (δPs), 190 internucleotide distances (D_{inter}) and 291 Residual Dipolar Coupling constants (RDCs: $\text{RDC}_{\text{C6/8-H6/8}}$, $\text{RDC}_{\text{C3'-H3'}}$, $\text{RDC}_{\text{C4'-H4'}}$ and $\text{RDC}_{\text{C1'-H1'}}$) were collected on the four dodecamers, discarding the first and last steps of each oligomer. This section examines the relations between the backbone linkage conformation and the relative orientations and distances between successive bases, using NMR data collected at 20°C.

Table 1. Sequences of the four studied DNA dodecamers constituents of the sequence 601.

Oligo 1	5'-TpCpGpTApGpCpApApGpCpT-3'•5'-ApGpCpTpTpGpCpTpApCpGpA-3'
Oligo 2	5'-GpCpTpCpTpApGpCpApCpCpG-3'•5'-CpGpGpTpGpCpTpApGpApGpC-3'
Oligo 3	5'-CpCpGpCpTpTpApApApCpGpC-3'•5'-GpCpGpTpTpTpApApGpCpGpG-3'
Oligo 4	5'-CpGpCpApCpGpTpApCpGpCpG-3'•5'-CpGpCpGpTpApCpGpTpGpCpG-3'

The δP values were extracted from well resolved ^1H - ^{31}P HETCOR spectra (Figure 2). They reflect the phosphate group dynamics that, in B-DNA, is dominated by the BI \leftrightarrow BII equilibrium that involves ϵ and ζ (Figure 1). Also, the presence of intense, well defined $^4J_{\text{H4'-P}}$ coupling constants testify that β/γ angles remain in the *trans/g+* region (see Materials and Methods), consistent with energy profiles (58,59).

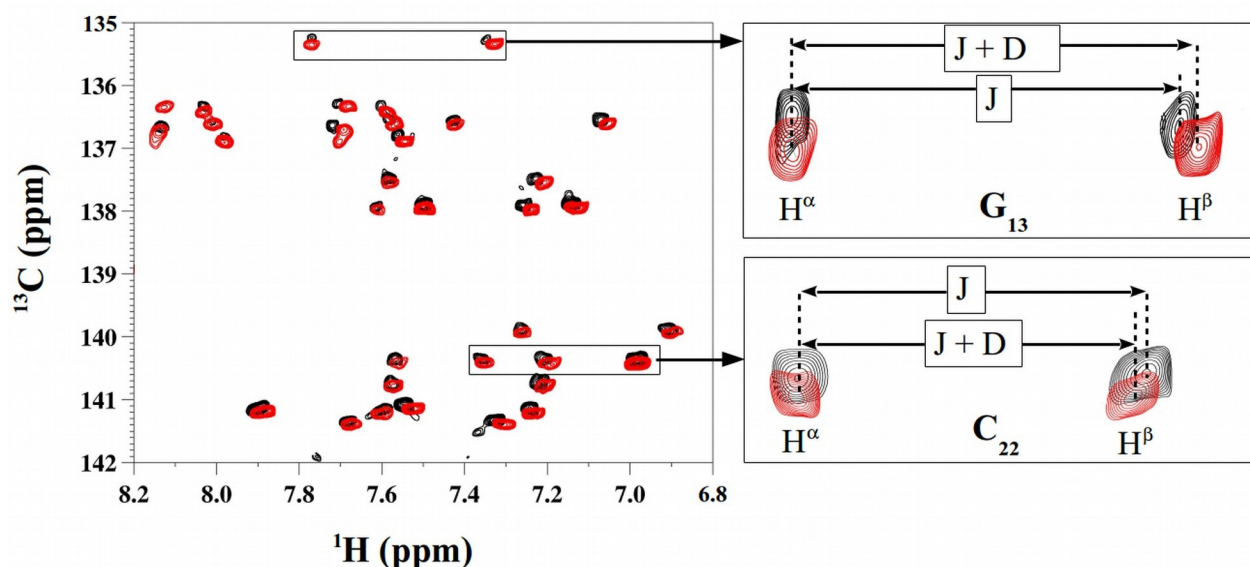
Figure 2 : Representative ^1H - ^{31}P HETCOR spectrum at 500 Mhz for δP determination.



The ^1H - ^{31}P HETCOR spectrum, used to determine δPs of pN residues (N for any base), shows a section of the H^4 - ^{31}P region collected with Oligo 2 at 20°C. The positive and negative cross-peaks are in black and red, respectively. The blue arrows indicate the cross-peak centers for specific bases. The terminal residues are indicated by stars.

For each type of RDC vector (see Figure 1), we examined ΔRDC , defined as the difference between the RDC values of two consecutive vectors on the same DNA strand, measured on residues i and $i+1$ ($\Delta\text{RDC} = \text{RDC}_{\text{residue } i+1} - \text{RDC}_{\text{residue } i}$). ΔRDC quantities contain information on the relative orientations of two successive vectors and therefore about the internal structure of a dinucleotide step, as δP and D_{inter} . Due to the global regularity of the double helix, a given family of RDC covers a more restricted range than those collected on proteins or RNA (60,61) (examples in Figure 3), limiting the number and precision of measurements (16). To evaluate the quality of our measurements, RDCs involving two types of C-H vectors belonging to a same base were compared. The vectors C6/8-H6/8 are coplanar with the C2-H2 vectors in adenines and C5-H5 in cytosines. Hence, the corresponding RDCs are expected to be very similar and may be used to test the quality of the measured RDCs. The correlation coefficients between $\text{RDC}_{\text{C6/8-H6/8}}$ and $\text{RDC}_{\text{C2-H2}}$ on one hand, and $\text{RDC}_{\text{C6/8-H6/8}}$ and $\text{RDC}_{\text{C5-H5}}$ on the other hand, are 0.95 and 0.89, respectively. This good agreement is reassuring, at least for the quality of RDCs involving these base atoms.

Figure 3 : Representative ^1H - ^{13}C TROSY spectra for RDC determination.



H^α H^β TROSY spectra obtained for Oligo 3 in isotropic solution (black peaks) and anisotropic medium (red peaks) were superimposed in the three panels presented here. The left panel represents a region of C6/8-H6/8 vectors. The right panels focus on two representative TROSY peaks involving G13 (top) and C22 (bottom). The splitting between H^α and H^β is indicated by arrows showing $^1J(\text{C-H})$ coupling in isotropic solution (J) and the sum of $^1J(\text{C-H})$ and residual dipolar coupling (RDC) in oriented medium ($J+D$). RDC is positive for G13 and negative for C22.

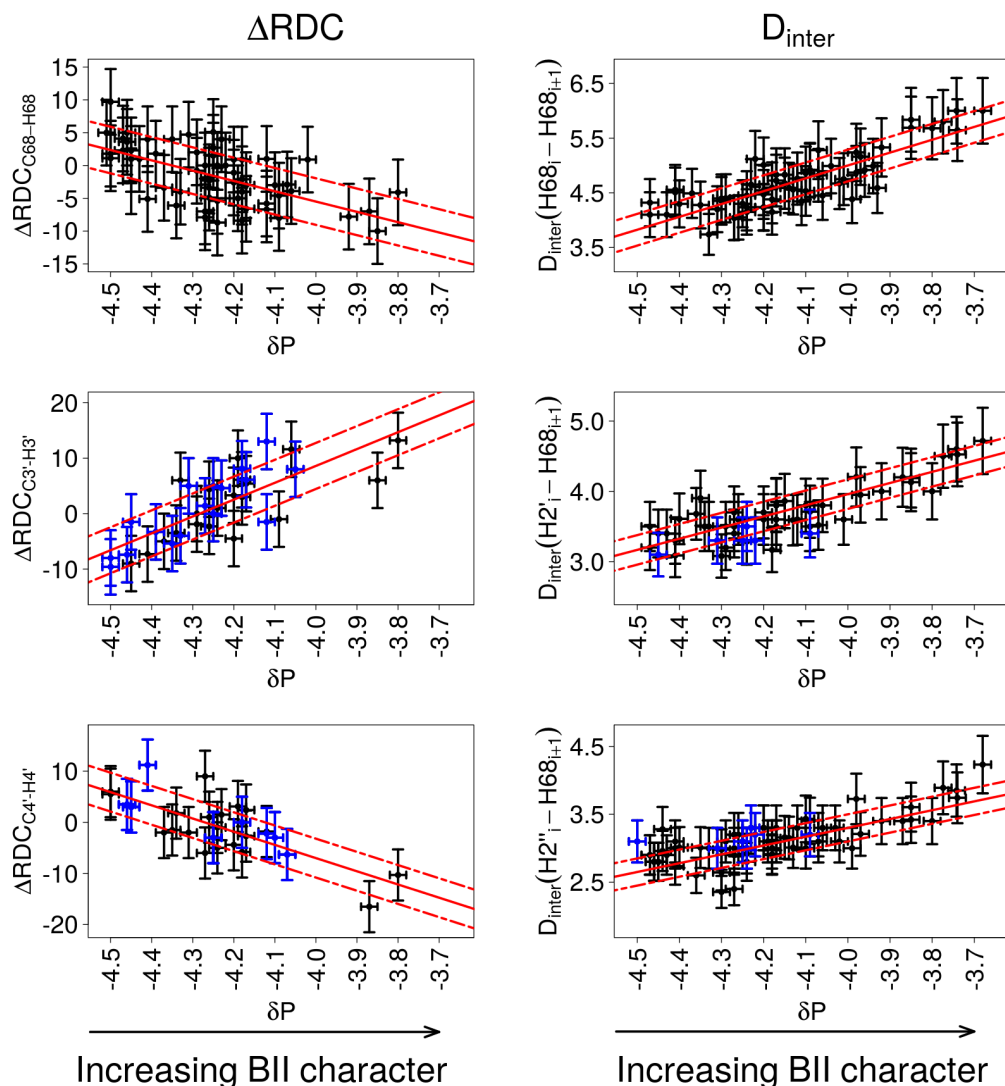
Notable correlations are found between three ΔRDCs ($\Delta\text{RDC}_{\text{C6/8-H6/8}}$, $\Delta\text{RDC}_{\text{C3'-H3'}}$ and $\Delta\text{RDC}_{\text{C4'-H4'}}$) and δPs (Figure 4 and Table 2; details per dodecamer in Table S1 in Supplementary data). The correlation coefficients (Table 2) provide evidence that those ΔRDCs are significantly coupled to the δPs . By comparison, the very low correlation coefficient of 0.2 between $\Delta\text{RDC}_{\text{C1'-H1'}}$ and δP shows that these quantities are not correlated. Distinctive trends characterize the couplings between ΔRDCs and δPs : $\Delta\text{RDC}_{\text{C6/8-H6/8}}$ and $\Delta\text{RDC}_{\text{C4'-H4'}}$ tend to be positive for high-field shifted δP (BI-rich step) and negative for down-field shifted δP (increased BII character), while the inverse is observed for $\Delta\text{RDC}_{\text{C3'-H3'}}$ (Figure 4).

Table 2: Relationships between δP s, ΔRDC s and internucleotide distances.

	N	CC	Linear fit	SD_{res}	p-value
ΔRDC_{C8-H8}	60	-0.60	$\Delta RDC_{C8-H8} = -15.7 \delta P - 68.3$	3.55	$5.44 \cdot 10^{-7}$
$\Delta RDC_{C3'-H3'}$	37	0.76	$\Delta RDC_{C3'-H3'} = 30.4 \delta P + 130.2$	4.15	$5.38 \cdot 10^{-8}$
$\Delta RDC_{C4'-H4'}$	29	-0.75	$\Delta RDC_{C4'-H4'} = -25.8 \delta P - 110.2$	3.8	$3.42 \cdot 10^{-6}$
$D_{inter}(H6/8_i-H6/8_{i+1})$	64	-0.81	$D_{inter}(H6/8_i-H6/8_{i+1}) = 2.3 \delta P + 14.3$	0.29	$< 2.2 \cdot 10^{-16}$
$D_{inter}(H2'_i-H6/8_{i+1})$	50	0.84	$D_{inter}(H2'_i-H6/8_{i+1}) = 1.58 \delta P + 10.2$	0.22	$6.20 \cdot 10^{-15}$
$D_{inter}(H2''_i-H6/8_{i+1})$	70	-0.76	$D_{inter}(H2''_i-H6/8_{i+1}) = 1.2 \delta P + 7.9$	0.21	$2.78 \cdot 10^{-14}$

This table reports the characteristics of the correlations shown in Figure 4 between δP s and i) ΔRDC s or ii) internucleotide distances (D_{inter}), based on NMR measurements at 20°C and excluding the terminal dinucleotides. The number of experimental data (N) and correlation coefficient (CC) are given for each type of ΔRDC s or D_{inter} . The equations correspond to linear fits of correlations, characterized here by the standard deviations of the normal distribution of the residuals (SD_{res}) and p-values. The dataset of D_{inter} contains additional measurements previously collected on the Jun-Fos oligomer (33,62). ΔRDC s are in Hz, δP in ppm and D_{inter} in Å.

Figure 4 : Correlations between δP and ΔRDC s or internucleotide distances.



The relative orientation between bases of a same strand in a base step is characterized either by ΔRDC s (Hz, left panels) or by internucleotide distances (\AA , right panels). The backbone conformation with respect to the $B I \leftrightarrow B II$ equilibrium is represented by its ^{31}P chemical shifts (δP , ppm). The NMR data were collected at 20°C on the four dodecamers studied here. The data exclude the terminal steps, which are shown in Figure 5.

Left: Correlations between ^{31}P chemical shifts and $\Delta RDC_{C6/8-H6/8}$, $\Delta RDC_{C3'-H3'}$ and $\Delta RDC_{C4'-H4'}$. ΔRDC s are defined as the difference between the RDC values of two consecutive vectors of a same type, measured on residues i and $i+1$: $\Delta RDC = RDC_{\text{residue } i+1} - RDC_{\text{residue } i}$.

Right: Correlations between ^{31}P chemical shifts and the internucleotide distances $D_{\text{inter}}(H6/8_i-H6/8_{i+1})$, $D_{\text{inter}}(H2'_i-H6/8_{i+1})$ and $D_{\text{inter}}(H2''_i-H6/8_{i+1})$. This dataset also contains measurements previously collected on the Jun-Fos oligomer (Heddi et al 2006). Note that $D_{\text{inter}}(H6/8_i-H6/8_{i+1}) > 5$

Å are extracted from NOESY spectra recorded with a mixing time of 300ms and are thus less precisely measured than shorter distances.

The points in blue in panels involving $\Delta RDC_{C3'-H3'}$, $\Delta RDC_{C4'-H4'}$, $D_{inter}(H2'_i-H6/8_{i+1})$ and $D_{inter}(H2''_i-H6/8_{i+1})$ correspond to dinucleotides with detectable populations of East or North sugars. The vertical bars are the estimated experimental errors on ΔRDC and D_{inter} (see Materials and Methods). The experimental error on δP (horizontal bars) is estimated to ± 0.05 ppm for the left panel plots and to ± 0.02 ppm for the right panel plots (see Materials and Methods). The best linear fits are represented with red lines; the standard deviations of the normal distribution of the residuals are illustrated by dashed red lines. The characteristics of the correlations are reported in Table 2.

The internucleotide distances $D_{inter}(H6/8_i-H6/8_{i+1})$, $D_{inter}(H2'_i-H6/8_{i+1})$ and $D_{inter}(H2''_i-H6/8_{i+1})$ are strongly correlated with δP (Figure 4, Table 2; details per dodecamer in Table S2 in Supplementary data). No coupling with δP is observed for $D_{inter}(H1'_i-H6/8_{i+1})$, an echo of the lack of correlation between δP and $\Delta RDC_{C1'-H1'}$. $D_{inter}(H6/8_i-H6/8_{i+1})$, $D_{inter}(H2'_i-H6/8_{i+1})$ and $D_{inter}(H2''_i-H6/8_{i+1})$ increase when δP is shifted down-field, when the BII population increases (Figure 4). That is consistent with the sequential distances involving H6/8 and the methyl group of thymines in NpT steps being always short (2.9 ± 0.2 Å on average) and their δP being high-field shifted (4.4 ± 0.1 ppm, BI-rich steps). The correlations between δP s and the three internucleotide distances, $D_{inter}(H6/8_i-H6/8_{i+1})$, $D_{inter}(H2'_i-H6/8_{i+1})$ and $D_{inter}(H2''_i-H6/8_{i+1})$, were previously observed on a 14bp DNA (33,62). That they are retrieved with different systems and a much larger dataset considerably strengthens the notion that such couplings represents a general property of free DNA in solution.

Four correlations involve NMR data on sugar protons, either 5' and 3' ($\Delta RDC_{C3'-H3'}$ and $\Delta RDC_{C4'-H4'}$), or only 5' ($D_{inter}(H2'_i-H6/8_{i+1})$ and $D_{inter}(H2''_i-H6/8_{i+1})$), of the phosphate linkage and could be sensitive to the sugar behavior. Investigations on sugar conformation (see Material and Methods) shows that *South* conformers largely prevail in purine nucleotides of the dodecamers, but observable TOCSY H1'-H4' cross-peaks indicate the presence of *North* or *East* conformers on 17 sugars of pyrimidine nucleotides (C7, C15 and C19 in Oligo 1; C2, C4, T5, C8, C10 and T16 in Oligo 2; C4, T6, C10 and T16 in Oligo 3; C9, C11, C15 and C23 in Oligo 4). Such feature is thus typically seen on cytosine sugars, in line with previous *ab initio* quantum mechanical predictions (63) and subsequent observations (9,25,42,64-67). Nevertheless, steps with 5'-sugars undergoing *South* \leftrightarrow *North* or *East* transitions do not stand out in the correlations shown in Figure 4.

The next section further examines the structural significance of the relations highlighted here, to better understand how they can be related to the intrinsic DNA mechanics.

ΔRDC s and internucleotide distances versus phosphate and sugar conformations

Here, our aim is to interpret structurally ΔRDC s and D_{inter} , using molecular models of the B-DNA helix. Ultrahigh resolution X-ray DNA structures (see Materials and Methods) were chosen as probably the best available templates to calculate $RDC_{C6/8-H6/8}$, $RDC_{C3'-H3'}$, $RDC_{C4'-H4'}$, their

corresponding ΔRDC s values and $D_{inter}(H6/8_i-H6/8_{i+1})$, $D_{inter}(H2'_i-H6/8_{i+1})$ and $D_{inter}(H2''_i-H6/8_{i+1})$ and to examine their relationship with first the phosphate and sugar conformations.

With the X-ray structures the backbone conformation is represented as two discrete states (BI or BII) defined from (ϵ - ζ) values instead of a continuous variable (δP) representing populations arising from the BI \leftrightarrow BII equilibrium. In our dataset, the sugars are overwhelmingly in the *South* conformation, with 24 nucleotides with *East* sugars, all associated to BI linkages, as previously reported (2,68). We first categorized the calculated ΔRDC s and D_{inter} according to BI and BII (Table 3). Then, the BI dinucleotides were divided in two sub-categories according to the *South* and *East* sugars. The average values of ΔRDC s and D_{inter} undoubtedly depend on BI and BII. 5'-*East* sugars do not significantly influence $D_{inter}(H2'_i-H6/8_{i+1})$ or $D_{inter}(H2''_i-H6/8_{i+1})$. $\Delta RDC_{C3'-H3'}$ and $\Delta RDC_{C4'-H4'}$ appear more sensitive to 5' or 3'-*East* sugars, but the large standard deviations preclude firm conclusions about this effect. Importantly, these results show a strong dependence of ΔRDC s and D_{inter} on the backbone state, BI or BII, qualitatively consistent with the relations shown in Figure 4.

Table 3: Values of ΔRDC s and internucleotide distances in BI and BII steps from X-ray structures.

	All N=108	BI N=70	BII N=38	BI and 5' and 3' <i>South</i> sugars N=48	BI and 5' or 3' <i>East</i> sugars N=22
ΔRDC_{C8-H8}	1.8 (5.9)	4.7 (4.1)	-3.7 (4.7)	-	-
$\Delta RDC_{C3'-H3'}$	-0.7 (6.5)	-3.2 (4.1)	3.9 (4.7)	-4.2 (5.9)	-1.3 (5.3)
$\Delta RDC_{C4'-H4'}$	1.4 (7.1)	5.4 (4.8)	-5.9 (4.2)	5.0 (4.6)	6.2 (5.2)
	All N=108	BI N=70	BII N=38	BI and 5' <i>South</i> sugar N=59	BI and 5' <i>East</i> sugar N=11
$D_{inter}(H6/8_i-H6/8_{i+1})$	5.2 (1.1)	4.6 (0.7)	6.3 (0.6)	-	-
$D_{inter}(H2'_i-H6/8_{i+1})$	3.8 (1.1)	3.1 (0.4)	5.1 (0.4)	3.1 (0.4)	2.9 (0.2)
$D_{inter}(H2''_i-H6/8_{i+1})$	2.9 (0.6)	2.6 (0.4)	3.5 (0.4)	2.6 (0.4)	3.0 (0.4)

Calculated ΔRDC (Hz) and internucleotide distances D_{inter} (Å) were obtained from 108 dinucleotide steps of the high resolution X-ray B-DNA structures listed in Materials and Methods. The average values of ΔRDC and D_{inter} are reported for all steps and then selecting either BI or BII steps. A second categorization was applied to the BI steps, according to the sugar conformation. Such partitioning is not applicable to the BII steps, which are exclusively surrounded by *South* sugars. *N* is the occurrence of each category. The standard deviations are in brackets.

Relations between DNA helicoidal parameters and ΔRDC s and the internucleotide distances

The relationship between the inter base parameters and either ΔRDC s or D_{inter} were then investigated, focusing on slide, roll and twist. These parameters are the most variable inter base-

pair parameters in our high-resolution X-ray dataset as well as in much larger datasets (1,2). Thus, they are especially good candidates when seeking meaningful couplings. In this section, the inter base parameter values are given for NpN dinucleotides, *i.e.* within one strand of the double helix, to be compatible with the derived-NMR Δ RDCs and D_{inter} .

The characteristics of the linear fits involving two variables, *i.e.* Δ RDC or D_{inter} on one hand and slide, roll or twist on the other hand, are summarized in Table 4.

Table 4 : Correlations between Δ RDCs or internucleotide distances and each inter base parameters of slide, roll and twist, from X-ray structures.

	Slide			Roll			Twist		
	CC	SD _{res}	p-value	CC	SD _{res}	p-value	CC	SD _{res}	p-value
Δ RDC _{C8-H8}	-0.62	0.49	$5.58 \cdot 10^{-13}$	0.70	6.1	$< 2.2 \cdot 10^{-16}$	-0.64	8.3	$1.53 \cdot 10^{-13}$
Δ RDC _{C3'-H3'}	0.40	0.58	$1.86 \cdot 10^{-5}$	-0.54	7.2	$2.12 \cdot 10^{-9}$	0.35	10.1	$1.94 \cdot 10^{-4}$
Δ RDC _{C4'-H4'}	-0.53	0.53	$2.49 \cdot 10^{-9}$	0.76	5.6	$< 2.2 \cdot 10^{-16}$	-0.62	8.5	$1.42 \cdot 10^{-12}$
$D_{\text{inter}}(\text{H6}/8_i\text{-H6}/8_{i+1})$	0.62	0.83	$1.22 \cdot 10^{-12}$	-0.82	4.9	$< 2.2 \cdot 10^{-16}$	0.84	5.9	$< 2.2 \cdot 10^{-16}$
$D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$	0.68	0.78	$1.27 \cdot 10^{-15}$	-0.89	3.9	$< 2.2 \cdot 10^{-16}$	0.88	5.1	$< 2.2 \cdot 10^{-16}$
$D_{\text{inter}}(\text{H2}''_i\text{-H6}/8_{i+1})$	0.84	0.35	$< 2.2 \cdot 10^{-16}$	-0.69	6.2	$< 2.2 \cdot 10^{-16}$	0.46	9.6	$4.82 \cdot 10^{-7}$

Calculated Δ RDC (Hz), internucleotide distances D_{inter} (Å), slide (Å), roll (°) and twist (°) were obtained from the high resolution X-ray B-DNA structures listed in Materials and Methods. The dataset contains a total of 108 dinucleotide steps.

The correlation coefficients (CC), standard deviations of the normal distribution of the residuals (SD_{res}) and p-values characterize the linear fits between either Δ RDC or D_{inter} and each of the three parameters of slide, roll and twist. Slide, roll and twist were calculated for each strand of the DNA double helix, to be consistent with Δ RDCs and D_{inter} .

In addition, multiple linear regressions were performed, which enable to find the equations optimized to express Δ RDC or D_{inter} as functions of the three combined parameters of slide, roll and twist (Table 5). In this latter approach, one should keep in mind that the slide variations cover ~ 2.5 Å while those of roll and twist correspond to $\sim 35^\circ$ (Figure 5); therefore only the coefficients associated to roll and twist can be directly compared. As expected, the correlation coefficients are higher with multiple linear regressions (Tables 4 and 5). However, the characteristics of the best bivariate fits of pairs of variables are also excellent (Table 4), suggesting that this simpler view provides relevant information. Indeed, both approaches lead to coherent results.

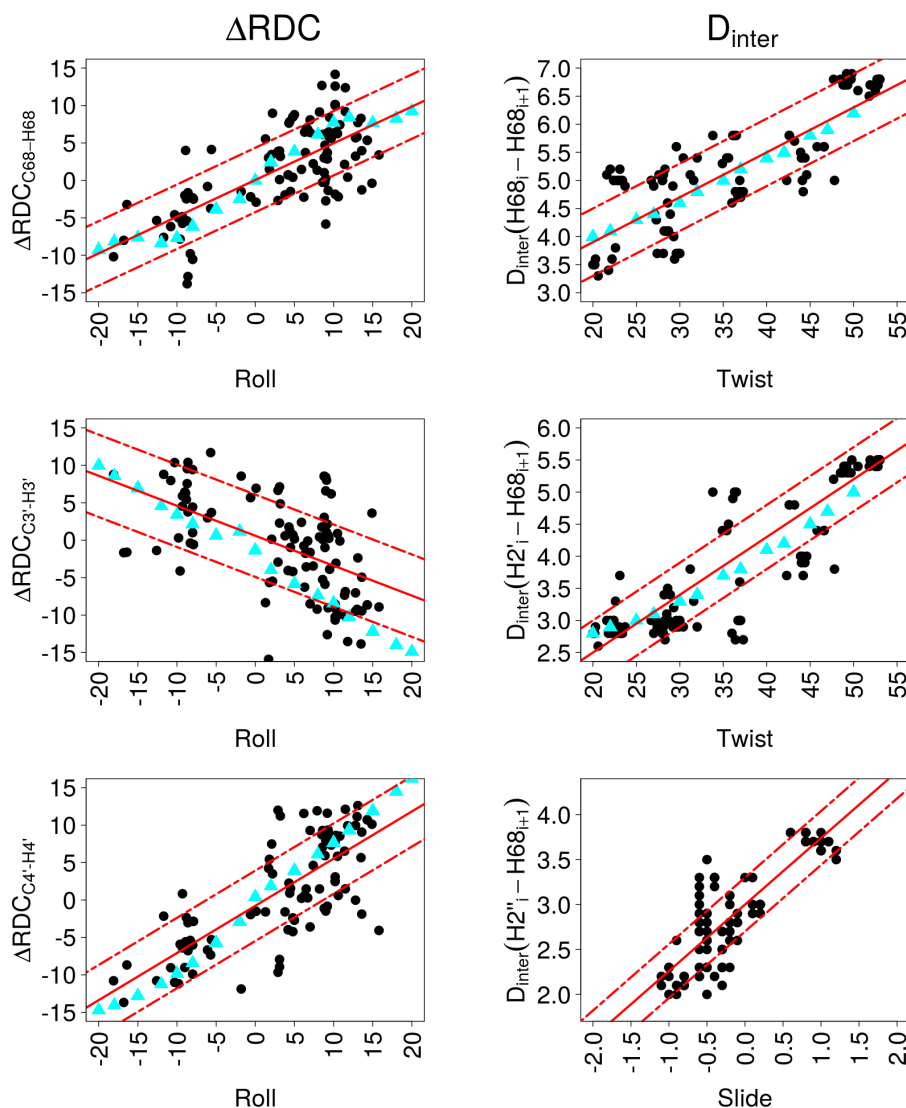
Table 5: Relationships between Δ RDCs or internucleotide distances and the combined inter base parameters of slide, roll and twist, from X-ray structures.

	CC	p-value
$\Delta RDC_{C8-H8} = -2.6 \text{ Slide} + 0.2 \text{ Roll} - 0.15 \text{ Twist} + 5.7$	0.91	$< 2.2 \cdot 10^{-16}$
$\Delta RDC_{C3'-H3'} = -0.06 \text{ Slide} - 0.53 \text{ Roll} - 0.12 \text{ Twist} + 5.1$	0.56	$3.6 \cdot 10^{-8}$
$\Delta RDC_{C4'-H4'} = 0.41 \text{ Slide} + 0.62 \text{ Roll} - 0.03 \text{ Twist} + 0.34$	0.75	$< 2.2 \cdot 10^{-16}$
$D_{\text{inter}}(H6/8_i-H6/8_{i+1}) = 0.2 \text{ Slide} - 0.04 \text{ Roll} + 0.05 \text{ Twist} + 3.5$	0.87	$< 2.2 \cdot 10^{-16}$
$D_{\text{inter}}(H2'_i-H6/8_{i+1}) = 0.2 \text{ Slide} - 0.05 \text{ Roll} + 0.05 \text{ Twist} + 2.3$	0.93	$< 2.2 \cdot 10^{-16}$
$D_{\text{inter}}(H2''_i-H6/8_{i+1}) = 0.6 \text{ Slide} - 0.02 \text{ Roll} + 0.005 \text{ Twist} + 3.2$	0.84	$< 2.2 \cdot 10^{-16}$

This table reports the equations expressing either $\Delta RDCs$ (Hz) or internucleotide distances (D_{inter} , Å) as function of the inter base parameters of slide (Å), roll (°) and twist (°); the equations were obtained from multiple linear regressions on data from high resolution X-ray structures. They are characterized by the correlation coefficients (CC) and the p-values.

$\Delta RDC_{C6/8-H6/8}$, $\Delta RDC_{C3'-H3'}$ and $\Delta RDC_{C4'-H4'}$ are more strongly coupled to roll than to twist or slide. In agreement with an earlier modeling study (69), $D_{\text{inter}}(H6/8_i-H6/8_{i+1})$ and $D_{\text{inter}}(H2'_i-H6/8_{i+1})$ are especially sensitive to roll and twist (Tables 4 and 5). $D_{\text{inter}}(H2''_i-H6/8_{i+1})$ appears as a good indicator of the slide value (Tables 4 and 5). Examples of the strongest couplings are illustrated in Figure 5. Despite their convincing characteristics, these correlations suffer from notable deviations that may just represent the natural variability in the present sample. They might also be a hint of biases which could affect the details of the X-ray DNA structures, such as lateral inter-molecular interactions (70-72) or underestimated structural heterogeneities (73).

Figure 5: Dependence of inter base parameters on ΔRDCs and D_{inter} .



These plots illustrate correlations between each of inter base parameters of roll, slide and twist versus residual dipolar coupling differences (ΔRDCs) or internucleotide distances (D_{inter}). ΔRDC (Hz), D_{inter} (Å), slide (Å), roll (°) and twist (°) were obtained from the high resolution X-ray B-DNA structures (black points, coefficient correlations in Table 4) or from 3DNA molecular models (cyan triangles). The best linear fits calculated for the X-ray structures correspond to the equations: $\text{Roll} = 1.0 \Delta\text{RDC}_{\text{C8-H8}} + 1.7$; $\text{Roll} = -0.7 \Delta\text{RDC}_{\text{C3'-H3'}} + 2.9$; $\text{Roll} = 0.9 \Delta\text{RDC}_{\text{C4'-H4'}} + 2.2$; $\text{Twist} = 8.5 D_{\text{inter}}(\text{H6/8i-H6/8i+1}) - 8.8$; $\text{Twist} = 9.0 D_{\text{inter}}(\text{H2'i-H6/8i+1}) + 1.3$; $\text{Slide} = 0.94 D_{\text{inter}}(\text{H2''i-H6/8i+1}) - 2.8$. They are represented by red lines; the standard deviations of the normal distribution of the residuals are traced by dashed red lines. The inter-base parameters were calculated for each strand of the DNA double helix, to be consistent with ΔRDCs and D_{inter} .

However, because of the couplings existing between roll and twist in B-DNA X-ray structures (with the present dataset: correlation coefficient of 0.8), and between roll and slide (with the present dataset: correlation coefficient of 0.75), it is difficult to interpret the relations in terms of the specific influence arising from each helicoidal parameter. To gain a better idea of the relative roles of these parameters, we built a series of simplified, but structurally controlled, models of B-DNA generated with 3DNA (56) (see Materials and Methods). These models contain one dinucleotide step in which the roll and twist were independently constrained (one parameter at a time) over the range of values observed in X-ray structures, keeping the other helical parameters at their canonical values (see Materials and methods). Δ RDCs and D_{inter} were then monitored on the constrained dinucleotide steps.

This approach reveals that neither twist nor slide variations affect Δ RDCs. Conversely, changes in roll generate Δ RDC variations that account for the Δ RDC versus roll correlations inferred from the X-ray structures (Figure 5). On the other hand, variations in roll or slide only produce minor changes in $D_{\text{inter}}(\text{H6}/8_i\text{-H6}/8_{i+1})$ and $D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$, compared to the increase of more than 2 Å when incrementing the twist from 20 to 55°. There is a remarkable agreement between the twist/ D_{inter} correlations obtained from the 3DNA models and the X-ray structures (Figure 5). Thus, Δ RDCs are primarily sensitive to the roll, while $D_{\text{inter}}(\text{H6}/8_i\text{-H6}/8_{i+1})$ and $D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$ are mainly modulated by the twist. The controlled models indicate that the correlations reported in Tables 4 and 5 also reflect the slide/roll and roll/twist couplings, as expected (see above); such redundancy should actually help when using the NMR observables for structure determination, making the refinement protocols more robust.

The relationships highlighted above first establish that both Δ RDCs and D_{inter} inform about inter base helicoidal parameters. Positive and negative values of Δ RDC_{C6/8-H6/8} and Δ RDC_{C4'-H4'} indicate positive and negative rolls, respectively, while the inverse is found for Δ RDC_{C3'-H3'}; both $D_{\text{inter}}(\text{H6}/8_i\text{-H6}/8_{i+1})$ and $D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$ increase with the twist value; small and large values of $D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$ reveal negative and positive slides, respectively. The couplings between δ Ps, Δ RDCs and D_{inter} (Figure 4) can now be firmly and coherently combined to reconstitute the intrinsic B-DNA mechanics, in which increasing BII populations are associated with more positive slide, more negative roll and higher twist.

Behavior of 3' and 5' terminal dinucleotides

In the above sections, only internal steps were considered. Indeed, the terminal dinucleotides are usually discarded from any DNA oligomer study in solution because of the fraying of the terminal base pairs that occurs on nano- and picosecond time scales (66,74). This fraying was recently described by molecular dynamics as very large motions that involve not only the terminal base pairs but also their neighbors (12,40,41). However, it is difficult to judge the realism of such motions since fraying events are very poorly documented by experiments. Thus, we turn to the NMR data collected on the outermost nucleotides and dinucleotides to appreciate to what extent they differ from the internal counterparts.

With NMR, monitoring the imino proton resonances in one-dimensional ^1H spectra is usually used to probe the stability of base pairing. More precisely, the absence of imino proton resonances reveals dynamic base pairs, weakly protected from proton exchange. For the four dodecamers, such spectra show that the imino proton resonances are lost in $\text{N}_1:\text{N}_{24}$, and $\text{N}_{12}:\text{N}_{13}$ (N standing for any base type) from 20°C , while they are clearly observable in all the other, internal base pairs, comprising $\text{N}_2:\text{N}_{23}$ and $\text{N}_{11}:\text{N}_{14}$. This is consistent with earlier NMR experiments (75,76) which indicated that the penultimate base pairs are much less subjected to long-lived opening events than the terminal base pairs. Despite the loss of Watson-Crick base pairing signature in $\text{N}_1:\text{N}_{24}$ and $\text{N}_{12}:\text{N}_{13}$, numerous NMR observables were collected on the outermost individual nucleotides, N_1 , N_{12} , N_{13} and N_{24} or dinucleotides, N_1pN_2 , $\text{N}_{11}\text{pN}_{12}$, $\text{N}_{13}\text{pN}_{14}$ and $\text{N}_{23}\text{pN}_{24}$.

At the nucleotide level, the intranucleotide distances $\text{H1}'_i\text{-H6}/8_i$ for the terminal nucleotides are on average $4.08 \pm 0.25 \text{ \AA}$ (values given in Table S3 in Supplementary data), typical of glycosidic torsion χ in the usual (47) and energetically most stable (77) *anti* orientation. The terminal sugars are flexible since their $\text{H1}'\text{-H4}'$ TROSY cross-peaks show signs of the of *East* or *North* conformers, except in T1 in Oligo 1 and G13 in Oligo 3 (Table S3 in Supplementary data).

Concerning the dinucleotides, the intense $^4J_{\text{H4}'\text{-P}}$ couplings seen in all 5'-terminal phosphates unambiguously testify of β/γ in *trans/g+* (these couplings cannot be observed on the 3' terminal step since they lack the corresponding phosphate). The δP s of both 3' and 5'-ends are very similar to those of internal phosphates, from 20° (Table 6; examples on Figure 2) to 40°C . The only exception was the terminal CpT of Oligo 1, down-field shifted by 0.25 ppm (gain of BII conformers) compared to their internal counterparts. It is intriguing that the sequence effect on δP s in B-DNAs, established from internal dinucleotides (11,35), appears to hold for these eight terminal steps (Table 6).

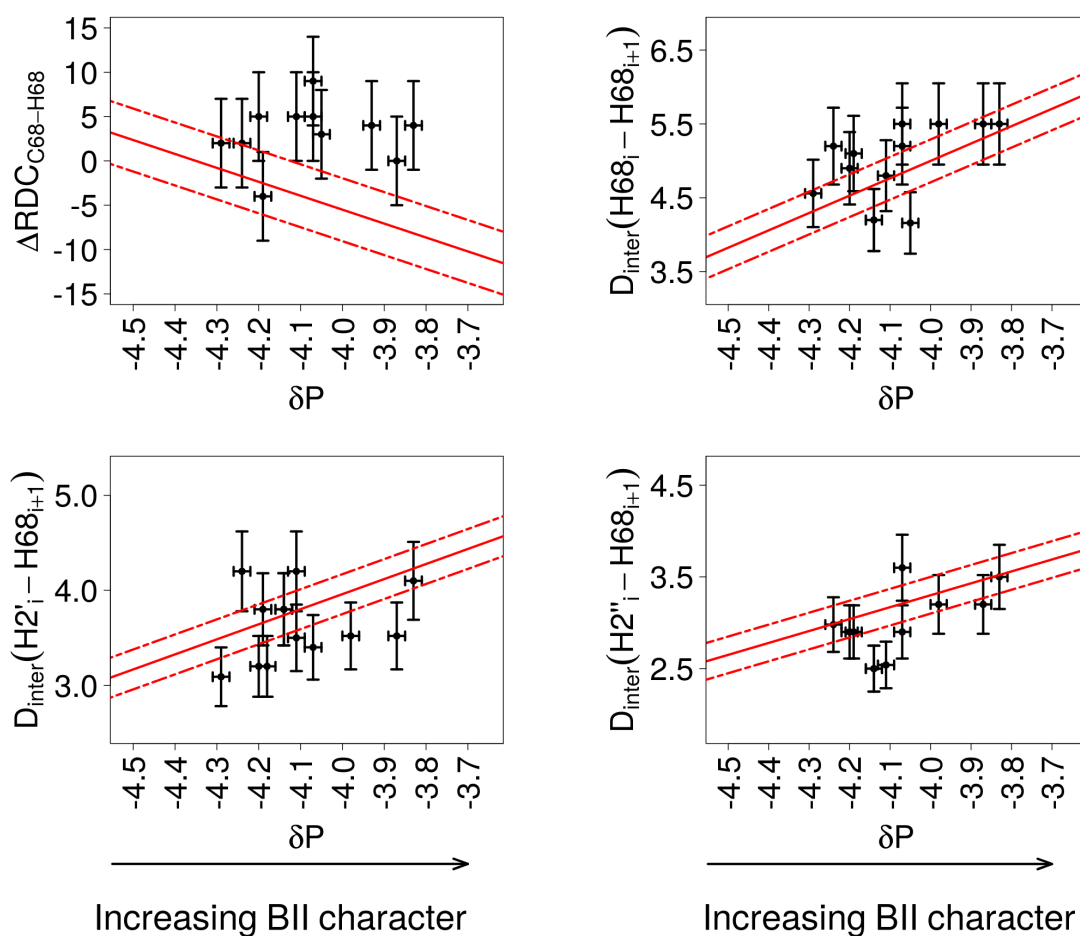
Table 6: Comparison between δP values of internal and terminal phosphates.

	δP of 5' or 3' ends	δP of internal steps	predicted δP
CpT	-4.18	-4.43 (0.07)	-4.38 (0.12)
TpC	-4.24	-4.25	-4.26 (0.12)
ApG	-4.20	-4.22 (0.03)	-4.21 (0.09)
GpC	-4.15 (0.11)	-4.14 (0.07)	-4.13 (0.08)
CpC	-4.11	-4.12	-4.07 (0.06)
GpA	-4.11	-4.02	-4.10 (0.09)
CpG	-4.01 (0.13)	-4.02 (0.13)	-4.02 (0.12)
GpG	-3.93	-3.85	-4.02 (0.09)

The δP values (in ppm) collected at 20°C are given for the first and last dinucleotide steps of the four dodecamers and for their internal counterparts. For dinucleotides present in several copies, δP is the average value, with the standard deviations in brackets. The last column reports δP s predicted depending on the dinucleotidic sequence (36).

We collected 14 $\text{RDC}_{\text{C6/8-H6/8}}$ (out of total of 16, see Table S3 in Supplementary data) on N_1 , N_{24} , N_{12} and N_{13} . $\Delta\text{RDC}_{\text{C6/8-H6/8}}$ involving either one terminal base and one internal base, or two internal bases, are in the same range and thus comparable. However, terminal $\Delta\text{RDC}_{\text{C6/8-H6/8}}$, in particular those associated to down-field shifted δP s, appear to depart from the correlation existing for the internal observables (Figure 6).

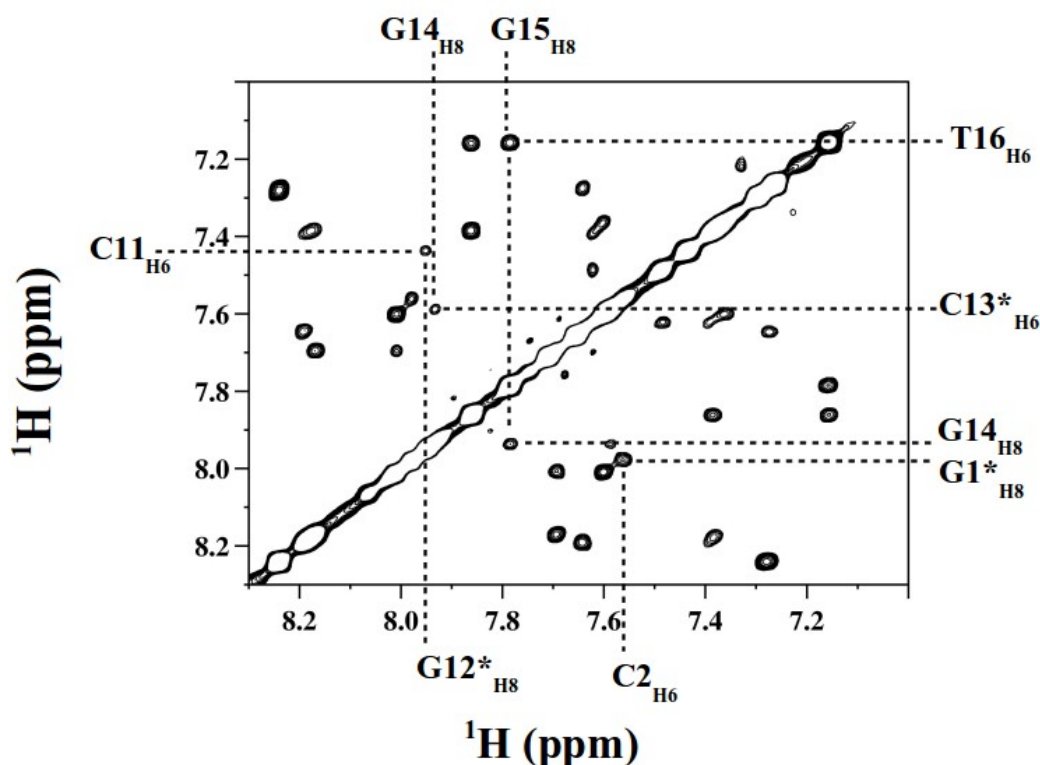
Figure 6: Behavior of terminal dinucleotides regarding δP versus ΔRDC s or internucleotide distances.



^{31}P chemical shifts (δP , ppm), $\Delta\text{RDC}_{\text{C6/8-H6/8}}$ (Hz), $D_{\text{inter}}(\text{H6/8}_i\text{-H6/8}_{i+1})$, $D_{\text{inter}}(\text{H2}'_i\text{-H6/8}_{i+1})$ and $D_{\text{inter}}(\text{H2}''_i\text{-H6/8}_{i+1})$ (\AA) were collected on the first and last dinucleotides of the four dodecamers, at 20°C . They are compared to the correlations obtained for the internal dinucleotides (see Figure 4), which are depicted here by the best linear fits (red lines) and the standard deviations of the normal distribution of the residuals (dashed red lines). The bars are the estimated experimental errors on ΔRDC and D_{inter} (vertical bars) and δP (horizontal bars).

Finally, one scrutinized the internucleotide distances (Table S4 in Supplementary data). $D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$ and $D_{\text{inter}}(\text{H2}''_i\text{-H6}/8_{i+1})$ were observed for all the terminal dinucleotides, in addition to 13 out of a total of 16 $D_{\text{inter}}(\text{H6}/8_i\text{-H6}/8_{i+1})$ (see examples in Figure 7).

Figure 7: Representative H8/6-H8/6 intra-strand connectivities on terminal and internal steps, from ^1H - ^1H NOESY spectrum.



The NOESY spectrum shows a section of a region of base-base protons collected with Oligo 2 at 20°C, using a mixing time of 300 ms. The crossing of dashed lines correspond to proton-proton connectivities. The terminal residues are indicated by stars. The connectivities involving terminal bases are as well observed as those involving internal bases.

Most of them were accurately measured from connectivities without overlaps. These distances tend to increase with down-field shifted δPs (Figure 6), but that may be coincidental rather than true correlations, considering the small sample sizes (10 to 12 points) and their variability. If these are weak correlations, they are much less convincing than those obtained with internal steps (Figure 4).

Thus, despite the absence of Watson-Crick hydrogen bonds at the oligonucleotide termini, the outermost dinucleotides maintain some characteristics of the double helix. RDC, ΔRDC and D_{inter} values indicate that the DNA does not fray into long-lived structures with large changes in the orientation of the terminal bases. Nevertheless, the correlations highlighted for internal base pairs are significantly weakened or even lost. This supports the intuitive perception that the terminal dinucleotides do not firmly adhere to the mechanics of a regular B-DNA helix, as analyzed in the above sections.

DISCUSSION

To build on previous studies (4,11,33,35), we explored here the dynamical structure of free B-DNA by NMR in solution. Our approach exploits data gathered on four B-DNA dodecamers, which contain a large variety of dinucleotide steps. We collected an exhaustive and consistent dataset including δP s, internucleotide distances and RDCs. Then, ΔRDC s were compiled as the difference between two sequentially consecutive RDCs on the same DNA strand.

First, this work highlights a series of couplings between NMR observables, some of them for the first time. The strong correlations between δP and the internucleotide distances $D_{\text{inter}}(\text{H6}/8_i\text{-H6}/8_{i+1})$, $D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$ and $D_{\text{inter}}(\text{H2}''_i\text{-H6}/8_{i+1})$, strengthen and generalize initial insights in this area reported for the 14bp Jun-Fos oligomer (33,62). New additional couplings are uncovered between δP and three ΔRDC s, $\Delta RDC_{\text{C6}/8\text{-H6}/8}$, $\Delta RDC_{\text{C3}'\text{-H3}'}$ and $\Delta RDC_{\text{C4}'\text{-H4}'}$. This shows that the RDCs can be exploited to gain a deeper understanding of the couplings between backbone states and orientation of the bases. In addition, our data suggest that the correlations involving either ΔRDC s or D_{inter} are not significantly affected by the sugar dynamics, presumably because the populations of *North* or *East* sugars remain modest. Incidentally, flexibility of the sugar was mainly discernible on cytosine residues, consistent with quantum mechanical calculations.

The mechanistic interpretation of the NMR-derived correlations involved calculating ΔRDC s and D_{inter} on ultrahigh resolution X-ray structures and 3DNA models, which enabled to relate these quantities to the structural descriptors of dinucleotides. In the X-ray structures, different ranges of ΔRDC s and D_{inter} are associated to the backbone BI and BII states. $\Delta RDC_{\text{C6}/8\text{-H6}/8}$ and D_{inter} do not seem influenced much by the discrete *South* and *East* sugar categories, accessible in the X-ray structures. That corroborates and strengthens the couplings observed from the NMR data. Importantly, our approach also reveals that i) ΔRDC s are especially sensitive to roll ii) $D_{\text{inter}}(\text{H6}/8_i\text{-H6}/8_{i+1})$ and $D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$ mostly depend on twist and iii) $D_{\text{inter}}(\text{H2}''_i\text{-H6}/8_{i+1})$ relate to slide. Therefore, ΔRDC s and D_{inter} collected by NMR can provide an estimate of three helical parameters, slide, roll and twist. Such estimates may not be very precise, at least for now, in view of the noise inevitably present in the underlying correlations. Also one must keep in mind that, in solution, the DNA is dynamic and thus the measured average values of ΔRDC s and D_{inter} inform about average values of the helical parameters.

The overall emerging trends offer a general framework for the intrinsic mechanics of DNA free in solution, generalizing and extending the earlier findings gleaned from X-ray data (2,3). Thus, BI-rich dinucleotides (high-field shifted δP) correlate with an average structure with negative slide (short $D_{\text{inter}}(\text{H2}''_i\text{-H6}/8_{i+1})$), null or positive roll (positive $\Delta RDC_{\text{C6}/8\text{-H6}/8}$, and $\Delta RDC_{\text{C4}'\text{-H4}'}$; negative $\Delta RDC_{\text{C3}'\text{-H3}'}$) and low twist (short $D_{\text{inter}}(\text{H6}/8_i\text{-H6}/8_{i+1})$ and $D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$); an increasing BII population (down-field shifted δP) is associated to shifts towards positive slide (large $D_{\text{inter}}(\text{H2}''_i\text{-H6}/8_{i+1})$), negative roll (negative $\Delta RDC_{\text{C6}/8\text{-H6}/8}$, and $\Delta RDC_{\text{C4}'\text{-H4}'}$; positive $\Delta RDC_{\text{C3}'\text{-H3}'}$) and high twist (large $D_{\text{inter}}(\text{H6}/8_i\text{-H6}/8_{i+1})$ and $D_{\text{inter}}(\text{H2}'_i\text{-H6}/8_{i+1})$).

We recall that δP and the associated BII propensities are primarily controlled at the dinucleotide level (36), comprising those of the four dodecamers studied here (11): CpG•CpG, CpA•TpG, GpC•GpC, GpG•CpC are characterized by BII percentages markedly higher than

average; ApN•NpT (N: any base) and TpA•TpA are globally BI-rich; GpA•TpC is an intermediate case, with BII percentage close to the average. According to our analysis, slide, roll and twist are expected to follow the same type of sequence dependence.

Finally, examining the NMR data collected on the oligomer termini shows that the correlations between δ Ps, Δ RDCs and D_{inter} are notably weakened or even lost for the 3' and 5' terminal steps and hence hold primarily for the internal nucleotides. However, the values of the NMR observables are in the same range for external and internal steps. This indicates that, on average, the outermost dinucleotides maintain some characteristics of the B-double helix and that the DNA fraying does not imply long-lived severe alteration of the orientation of the terminal bases.

Overall, this work provides new information on the structural significance of a range of NMR observables for DNA. In addition, it offers suitable reference data against which MD-simulated DNA structures may be compared, with or without NMR restraints. The number of long unrestrained MD simulations of DNA has grown quickly in recent years (78-82), however the development and tests of the underlying force-fields are hampered by the dearth of detailed experimental structural information for DNA in solution. An illustration of one probable MD shortcoming is the behavior of the DNA termini. In recent simulations (40, 41, 82), the stacking between terminal bases and neighbors was disrupted, with dramatic changes of χ and backbone angles and emergence of various interactions involving the bases, including contacts with internal phosphate groups, insertion into the minor groove or mispairing with internal base. Such motions and interactions, which incidentally compromise the convergence of the simulations, are suspected to be artifacts (41), a view compatible with the present NMR data. In such context, the results from the study of the four oligomers should help assess and improve the representation of DNA by force-fields.

It is now widely accepted that the sequence specific properties of free B-DNA, relevant for instance to its indirect readout, packaging and recognition, are mediated by significant but subtle local conformational heterogeneity. This heterogeneity has historically been difficult to detect and characterize in solution because it is dynamic and involves only limited departures from the overall helical pattern. The couplings documented here provide new evidence of the intrinsic DNA mechanics, which corresponds to the tight relationship between backbone and base behavior. Importantly, each type of dinucleotide explores the BI and BII profiles differently, resulting in variable deformability along the double helix. This strengthens the notion that the TRX scale, which summarizes this sequence dependence, provides a fundamental guide to understand more completely the fine local structure of B-DNA, and how it modulates its recognition by proteins.

ACKNOWLEDGEMENTS

The East China Normal University is gratefully acknowledged for its support.

REFERENCES

1. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. and Zhurkin, V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 11163-11168.
2. Djuranovic, D. and Hartmann, B. (2003) Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *Journal of biomolecular structure & dynamics*, **20**, 771-788.
3. Djuranovic, D. and Hartmann, B. (2004) DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers*, **73**, 356-368.
4. Oguey, C., Foloppe, N. and Hartmann, B. (2010) Understanding the sequence-dependence of DNA groove dimensions: implications for DNA interactions. *PLoS One*, **5**, e15931.
5. Prive, G.G., Heinemann, U., Chandrasegaran, S., Kan, L.S., Kopka, M.L. and Dickerson, R.E. (1987) Helix geometry, hydration, and G.A mismatch in a B-DNA decamer. *Science*, **238**, 498-504.
6. Prive, G.G., Yanagi, K. and Dickerson, R.E. (1991) Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. *Journal of molecular biology*, **217**, 177-199.
7. Schneider, B., Neidle, S. and Berman, H.M. (1997) Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers*, **42**, 113-124.
8. van Dam, L. and Levitt, M.H. (2000) BII nucleotides in the B and C forms of natural-sequence polymeric DNA: A new model for the C form of DNA. *Journal of molecular biology*, **304**, 541-561.
9. van Dam, L., Ouwerkerk, N., Brinkmann, A., Raap, J. and Levitt, M.H. (2002) Solid-state NMR determination of sugar ring pucker in (13)C-labeled 2'-deoxynucleosides. *Biophysical journal*, **83**, 2835-2844.
10. Fratini, A.V., Kopka, M.L., Drew, H.R. and Dickerson, R.E. (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J Biol Chem*, **257**, 14686-14707.
11. Xu, X., Ben Imeddourene, A., Zargarian, L., Foloppe, N., Mauffret, O. and Hartmann, B. (2014) NMR studies of DNA support the role of pre-existing minor groove variations in nucleosome indirect readout. *Biochemistry*, **53**, 5601-5612.
12. Drsata, T., Perez, A., Orozco, M., Morozov, A.V., Sponer, J. and Lankas, F. (2013) Structure, Stiffness and Substates of the Dickerson-Drew Dodecamer. *J Chem Theory Comput*, **9**, 707-721.
13. Hartmann, B., Piazzola, D. and Lavery, R. (1993) BI-BII transitions in B-DNA. *Nucleic acids research*, **21**, 561-568.
14. Winger, R.H., Liedl, K.R., Pichler, A., Hallbrucker, A. and Mayer, E. (1999) Helix morphology changes in B-DNA induced by spontaneous B(I) \rightleftharpoons B(II) substrate interconversion. *Journal of biomolecular structure & dynamics*, **17**, 223-235.
15. Zacharias, M. (2006) Minor groove deformability of DNA: a molecular dynamics free energy simulation study. *Biophysical journal*, **91**, 882-891.

16. Alvarez-Salgado, F., Desvaux, H. and Boulard, Y. (2006) NMR assessment of the global shape of a non-labelled DNA dodecamer containing a tandem of G-T mismatches. *Magn Reson Chem*, **44**, 1081-1089.
17. McAteer, K., Aceves-Gaona, A., Michalczyk, R., Buchko, G.W., Isern, N.G., Silks, L.A., Miller, J.H. and Kennedy, M.A. (2004) Compensating bends in a 16-base-pair DNA oligomer containing a T(3)A(3) segment: A NMR study of global DNA curvature. *Biopolymers*, **75**, 497-511.
18. Renisio, J.G., Cosquer, S., Cherrak, I., El Antri, S., Mauffret, O. and Femandjian, S. (2005) Pre-organized structure of viral DNA at the binding-processing site of HIV-1 integrase. *Nucleic acids research*, **33**, 1970-1981.
19. Vermeulen, A., Zhou, H. and Pardi, A. (2000) Determining DNA Global Structure and DNA Bending by Application of NMR Residual Dipolar Couplings. *Journal of the American Chemical Society*, **122**, 9638-9647.
20. Lipsitz, R.S. and Tjandra, N. (2004) Residual dipolar couplings in NMR structure analysis. *Annu Rev Biophys Biomol Struct*, **33**, 387-413.
21. MacDonald, D. and Lu, P. (2002) Residual dipolar couplings in nucleic acid structure determination. *Current opinion in structural biology*, **12**, 337-343.
22. Bae, S.H., Cheong, H.K., Cheong, C., Kang, S., Hwang, D.S. and Choi, B.S. (2003) Structure and dynamics of hemimethylated GATC sites: implications for DNA-SeqA recognition. *J Biol Chem*, **278**, 45987-45993.
23. Barbic, A., Zimmer, D.P. and Crothers, D.M. (2003) Structural origins of adenine-tract bending. *Proceedings of the National Academy of Sciences*, **100**, 2369-2373.
24. MacDonald, D., Herbert, K., Zhang, X., Polgruto, T. and Lu, P. (2001) Solution structure of an A-tract DNA bend. *Journal of molecular biology*, **306**, 1081-1098.
25. Schwieters, C.D. and Clore, G.M. (2007) A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry*, **46**, 1152-1166.
26. Tjandra, N., Tate, S.-i., Ono, A., Kainosho, M. and Bax, A. (2000) The NMR Structure of a DNA Dodecamer in an Aqueous Dilute Liquid Crystalline Phase. *Journal of the American Chemical Society*, **122**, 6190-6200.
27. Wu, Z., Delaglio, F., Tjandra, N., Zhurkin, V.B. and Bax, A. (2003) Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and ³¹P chemical shift anisotropy. *Journal of biomolecular NMR*, **26**, 297-315.
28. Stefl, R., Wu, H., Ravindranathan, S., Sklenar, V. and Feigon, J. (2004) DNA A-tract bending in three dimensions: Solving the dA4T4 vs. dT4A4 conundrum. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 1177-1182.
29. Gorenstein, D.G. (1992) ³¹P NMR of DNA. *Methods Enzymol*, **211**, 254-286.
30. Gorenstein, D.G. (1994) Conformation and Dynamics of DNA and Protein-DNA Complexes by ³¹P NMR. *Chemical Review*, **94**, 1315-1338.
31. Precechtelova, J., Munzarova, M.L., Vaara, J., Novotny, J., Dracinski, M. and Sklenar, V. (2013) Toward Reproducing Sequence Trends in Phosphorus Chemical Shifts for Nucleic

- Acids by MD/DFT Calculations. *Journal of Chemical Theory and Computation*, **9**, 1641-1656.
32. Precechtelova, J., Novak, P., Munzarova, M.L., Kaupp, M. and Sklenar, V. (2010) Phosphorus Chemical Shifts in a Nucleic Acid Backbone from Combined Molecular Dynamics and Density Functional Calculations. *Journal of the American Chemical Society*.
 33. Heddi, B., Foloppe, N., Bouchemal, N., Hantz, E. and Hartmann, B. (2006) Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition. *Journal of the American Chemical Society*, **128**, 9170-9177.
 34. Tian, Y., Kayatta, M., Shultis, K., Gonzalez, A., Mueller, L.J. and Hatcher, M.E. (2009) (31)P NMR Investigation of Backbone Dynamics in DNA Binding Sites (dagger). *J Phys Chem B*, **113**, 2596-2603.
 35. Heddi, B., Abi-Ghanem, J., Lavigne, M. and Hartmann, B. (2010) Sequence-dependent DNA flexibility mediates DNase I cleavage. *Journal of molecular biology*, **395**, 123-133.
 36. Heddi, B., Oguey, C., Lavelle, C., Foloppe, N. and Hartmann, B. (2010) Intrinsic flexibility of B-DNA: the experimental TRX scale. *Nucleic acids research*, **38**, 1034-1047.
 37. Thastrom, A., Bingham, L.M. and Widom, J. (2004) Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. *Journal of molecular biology*, **338**, 695-709.
 38. Meiler, J., Blomberg, N., Nilges, M. and Griesinger, C. (2000) A new approach for applying residual dipolar couplings as restraints in structure elucidation. *Journal of biomolecular NMR*, **16**, 245-252.
 39. van Dijk, A.D.J., Fushman, D. and Bonvin, A.M.J.J. (2005) Various strategies of using residual dipolar couplings in NMR-driven protein docking: Application to Lys48-linked di-ubiquitin and validation against 15N-relaxation data. *Proteins: Structure, Function, and Bioinformatics*, **60**, 367-381.
 40. Galindo-Murillo, R., Roe, D.R. and Cheatham, T.E., 3rd. (2014) Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). *Biochim Biophys Acta*, **1850**, 1041-1058.
 41. Zgarbova, M., Otyepka, M., Sponer, J., Lankas, F. and Jurecka, P. (2014) Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *Journal of Chemical Theory and Computation*, **10**, 3177-3189.
 42. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR*, **6**, 277-293.
 43. Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of biomolecular NMR*, **2**, 661-665.
 44. Baleja, J.D., Moulton, J. and Sykes, B.D. (1990) Distance Measurement and Structure Refinement with Noe Data. *Journal of Magnetic Resonance*, **87**, 375-384.
 45. Sklenar, V., Miyashiro, H., Zon, G., Miles, H.T. and Bax, A. (1986) Assignment of the 31P and 1H resonances in oligonucleotides by two-dimensional NMR spectroscopy. *FEBS Lett*, **208**, 94-98.

46. Wu, Z., Tjandra, N. and Bax, A. (2001) ^{31}P chemical shift anisotropy as an aid in determining nucleic acid structure in liquid crystals. *Journal of the American Chemical Society*, **123**, 3617-3618.
47. Wijmenga, S.S. and van Buuren, B.N.M. (1998) The use of NMR methods for conformational studies of nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **32**, 287-387.
48. Yang, J., McAteer, K., Silks, L.A., Wu, R., Isern, N.G., Unkefer, C.J. and Kennedy, M.A. (2000) A Comprehensive Approach for Accurate Measurement of Proton-Proton Coupling Constants in the Sugar Ring of DNA. *Journal of Magnetic Resonance*, **146**, 260-276.
49. Kim, S.G. and Reid, B.R. (1992) Solution structure of the TnAn DNA duplex GCCGTTAACGCG containing the HpaI restriction site. *Biochemistry*, **31**, 12103-12116.
50. Hansen, M.R., Hanson, P. and Pardi, A. (2000) Pfl filamentous phage as an alignment tool for generating local and global structural information in nucleic acids. *Journal of biomolecular structure & dynamics*, **17 Suppl 1**, 365-369.
51. Hansen, M.R., Mueller, L. and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol*, **5**, 1065-1074.
52. Hansen, M.R., Hanson, P. and Pardi, A. (2000) Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol*, **317**, 220-240.
53. Brutscher, B., Boisbouvier, J.-M., Pardi, A., Marion, D. and Simorre, J.-P. (1998) Improved Sensitivity and Resolution in ^1H - ^{13}C NMR Experiments of RNA. *Journal of the American Chemical Society*, **120**, 11845-11851.
54. Brutscher, B., Skrynnikov, N.R., Bremi, T., Bryschweiler, R. and Ernst, R.R. (1998) Quantitative Investigation of Dipole-CSA Cross-Correlated Relaxation by ZQ/DQ Spectroscopy. *Journal of Magnetic Resonance*, **130**, 346-351.
55. Zweckstetter, M. (2008) NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc*, **3**, 679-690.
56. Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic acids research*, **31**, 5108-5121.
57. Lavery, R. and Sklenar, H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *Journal of biomolecular structure & dynamics*, **6**, 63-91.
58. Foloppe, N. and MacKerell, A.D. (1999) Contribution of the Phosphodiester Backbone and Glycosyl Linkage Intrinsic Torsional Energetics to DNA Structure and Dynamics. *The Journal of Physical Chemistry B*, **103**, 10955-10964.
59. Varnai, P., Djuranovic, D., Lavery, R. and Hartmann, B. (2002) Alpha/gamma transitions in the B-DNA backbone. *Nucleic acids research*, **30**, 5398-5406.
60. Clore, G.M., Gronenborn, A.M. and Bax, A. (1998) A Robust Method for Determining the Magnitude of the Fully Asymmetric Alignment Tensor of Oriented Macromolecules in the Absence of Structural Information. *Journal of Magnetic Resonance*, **133**, 216-221.

61. Walsh, J.D., Cabello-Villegas, J. and Wang, Y.-X. (2004) Periodicity in Residual Dipolar Couplings and Nucleic Acid Structures. *Journal of the American Chemical Society*, **126**, 1938-1939.
62. Heddi, B., Foloppe, N., Hantz, E. and Hartmann, B. (2007) The DNA structure responds differently to physiological concentrations of K(+) or Na(+). *Journal of molecular biology*, **368**, 1403-1411.
63. Foloppe, N. and MacKerell, A.D., Jr. (1999) Intrinsic conformational properties of deoxyribonucleosides: implicated role for cytosine in the equilibrium among the A, B, and Z forms of DNA. *Biophysical journal*, **76**, 3206-3218.
64. Duchardt, E., Nilsson, L. and Schleucher, J. (2008) Cytosine ribose flexibility in DNA: a combined NMR ¹³C spin relaxation and molecular dynamics simulation study. *Nucleic acids research*, **36**, 4211-4219.
65. Gorler, A., Ulyanov, N.B. and James, T.L. (2000) Determination of the populations and structures of multiple conformers in an ensemble from NMR data: multiple-copy refinement of nucleic acid structures using floating weights. *Journal of biomolecular NMR*, **16**, 147-164.
66. Nikolova, E.N., Bascom, G.D., Andricioaei, I. and Al-Hashimi, H.M. (2012) Probing sequence-specific DNA flexibility in a-tracts and pyrimidine-purine steps by nuclear magnetic resonance (¹³C) relaxation and molecular dynamics simulations. *Biochemistry*, **51**, 8654-8664.
67. Shajani, Z. and Varani, G. (2008) ¹³C relaxation studies of the DNA target sequence for hhai methyltransferase reveal unique motional properties. *Biochemistry*, **47**, 7617-7625.
68. Isaacs, R.J. and Spielmann, H.P. (2001) NMR evidence for mechanical coupling of phosphate B(I)-B(II) transitions with deoxyribose conformational exchange in DNA. *Journal of molecular biology*, **311**, 149-160.
69. Lefebvre, A., Femandjian, S. and Hartmann, B. (1997) Sensitivity of NMR internucleotide distances to B-DNA conformation: underlying mechanics. *Nucleic acids research*, **25**, 3855-3862.
70. Johansson, E., Parkinson, G. and Neidle, S. (2000) A new crystal form for the dodecamer C-G-C-G-A-A-T-T-C-G-C-G: symmetry effects on sequence-dependent DNA structure. *Journal of molecular biology*, **300**, 551-561.
71. Minasov, G., Tereshko, V. and Egli, M. (1999) Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing. *Journal of molecular biology*, **291**, 83-99.
72. Wynveen, A., Lee, D.J., Kornyshev, A.A. and Leikin, S. (2008) Helical coherence of DNA in crystals and solution. *Nucleic acids research*, **36**, 5540-5551.
73. Maehigashi, T., Hsiao, C., Woods, K.K., Moulaei, T., Hud, N.V. and Williams, L.D. (2011) B-DNA structure is intrinsically polymorphic: even at the level of base pair positions. *Nucleic acids research*, **40**, 3714-3722.
74. Nikolova, E. and Al-Hashimi, H. (2009) Preparation, resonance assignment, and preliminary dynamics characterization of residue specific ¹³C/¹⁵N-labeled elongated DNA

- for the study of sequence-directed dynamics by NMR. *Journal of biomolecular NMR*, **45**, 9-16.
75. Leroy, J.L., Kochoyan, M., Huynh-Dinh, T. and Gueron, M. (1988) Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *Journal of molecular biology*, **200**, 223-238.
 76. Nonin, S., Leroy, J.L. and Gueron, M. (1995) Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. *Biochemistry*, **34**, 10652-10659.
 77. Foloppe, N., Hartmann, B., Nilsson, L. and MacKerell, A.D., Jr. (2002) Intrinsic conformational energetics associated with the glycosyl torsion in DNA: a quantum mechanical study. *Biophysical journal*, **82**, 1554-1569.
 78. Hart, K., Foloppe, N., Baker, C.M., Denning, E.J., Nilsson, L. and Mackerell, A.D., Jr. (2012) Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J Chem Theory Comput*, **8**, 348-362.
 79. Perez, A., Lankas, F., Luque, F.J. and Orozco, M. (2008) Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic acids research*, **36**, 2379-2394.
 80. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T.E., 3rd, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophysical journal*, **92**, 3817-3829.
 81. Vanommeslaeghe, K. and MacKerell, A.D., Jr. (2015) CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim Biophys Acta*, **1850**, 861-871.
 82. Zgarbova, M., Luque, F.J., Sponer, J., Cheatham, T.E., 3rd, Otyepka, M. and Jurecka, P. (2013) Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J Chem Theory Comput*, **9**, 2339-2354.

SUPPLEMENTARY DATA

Table S1 : Correlations between $\Delta RDCs$ and δPs for each of the four dodecamers

	$\Delta RDC_{C8-H8}/\delta P$		$\Delta RDC_{C3'-H3'}/\delta P$		$\Delta RDC_{C4'-H4'}/\delta P$	
	N	CC	N	CC	N	CC
Oligo 1	16	-0.83	12	0.76	7	-0.79
Oligo 2	18	-0.52	11	0.62	9	-0.84
Oligo 3	18	-0.61	11	0.95	11	-0.84
Oligo 4	8	-0.93	3	--	2	--

This table reports the characteristics of the correlations between $\Delta RDCs$ and δPs , obtained from the NMR measurements. These correlations are illustrated in Figure 4. The number of experimental data (N) and correlation coefficient (CC) for each dodecamers is given for each type of $\Delta RDCs$.

Table S2 : Correlations between D_{inter} and δPs for each of the four dodecamers

	$D_{inter}(H6/8_i-H6/8_{i+1})/\delta P$		$D_{inter}(H2'_i-H6/8_{i+1})/\delta P$		$D_{inter}(H2''_i-H6/8_{i+1})/\delta P$	
	N	CC	N	CC	N	CC
Oligo 1	8	0.76	10	0.89	11	0.82
Oligo 2	7	0.71	7	0.83	8	0.63
Oligo 3	6	0.68	9	0.91	14	0.77
Oligo 4	9	0.73	7	0.53	9	0.72

This table reports the characteristics of the correlations between D_{inter} and δPs , obtained from the NMR measurements. These correlations are illustrated in Figure 4. The number of experimental data (N) and correlation coefficient (CC) for each dodecamers is given for each type of D_{inter} .

Table S3: NMR data collected on terminal nucleotides

		$D_{\text{intra}}(\text{H1}'\text{-H6}/8)$	$I(\text{TOCSY}_{\text{H1}'\text{-H4}'})$	$\text{RDC}_{\text{C6}/8\text{-H6}/8}$
Oligo1	T1	4.5	0	12
	T12	3.75	2.22e+07	Overlap
	A13	4.3	6.40e+06	5
	A24	4.22	7.22e+06	Overlap
Oligo2	G1	4.5	3.54e+05	12
	G12	4.02	1.39e+06	7
	C13	Overlap	6.90e+05	7
	C24	4.02	2.61e+06	3
Oligo3	C1	4.3	8.19e+07	9
	C12	3.9	8.70e+07	6
	G13	4.2	0	12
	G24	4.1	1.47e+08	8
Oligo4	C1	3.8	1.49e+08	6
	G12	3.9	1.68e+08	7
	C13	3.8	1.49e+08	6
	G24	3.9	1.68e+08	7

The $D_{\text{intra}}(\text{H1}'\text{-H6}/8)$ values (in Å) reported here are typical of the anti orientation of the glycosidic bond. The intensity of H1'-H4' TROSY cross-peaks, null and thus not observable for pure South sugars, indicates a population of North or East on most terminal sugars. The $\text{RDC}_{\text{C6}/8\text{-H6}/8}$ values (in Hz) are similar to those obtained on internal nucleotides. The data were collected at 20°C. Some of them are observable but not measurable because of overlaps.

Table S4: Interdinucleotide distances collected on terminal dinucleotides

		D_{inter} (H2' _i -H6/8 _{i+1})	D_{inter} (H2'' _i -H6/8 _{i+1})	D_{inter} (H6/8 _i -H6/8 _{i+1})	D_{inter} (H5 _i -H6/8 _{i+1}) or (H6/8 _i -H5 _{i+1})
Oligo 1	T1pC2	4.2	2.58	5.2	5.2
	C11pT12	3.2	Overlap	-	-
	A13pG14	3.2	2.5	4.9	NO
	G23pA24	4.2	Overlap	-	NO
Oligo 2	G1pC2	3.8	2.5	5.1	~5.5
	C11pG12	4.1	3.1	~5.5	~5.5
	C13pG14	3.4	3.2	~5.5	-
	G23pC24	Overlap	Overlap	4.16	4.16
Oligo 3	C1pC2	3.5	2.14	4.8	4.6
	G11pC12	Overlap	2.5	5.2	4.3
	G13pC14	3.09	Overlap	4.56	4.02
	G23pG24	Overlap	Overlap	-	NO
Oligo 4	C1pG2	3.8	2.1	4.2	4.3
	C11pG12	3.52	2.8	~5.5	~5.5
	C13pG14	3.8	2.1	4.2	4.3
	C23pG24	3.52	2.8	~5.5	~5.5

$D_{inter}(H2'_i-H6/8_{i+1})$ and $D_{inter}(H2''_i-H6/8_{i+1})$ are observable on all the terminal dinucleotides. D_{inter} only observable in NOESY spectra recorded with a mixing time of 300ms were estimated to ~5.5Å. Some D_{inter} involving base protons are either not observed (-) or observed but not measured

because of overlaps. NO : not observable, such as internucleotides distances involving the H5 atom of cytosine in steps devoid of cytosine. The data were collected at 20°C.

CHAPITRE 4

Mouvements lents de l'ADN-B par RMN

1 Introduction générale

Dans ce chapitre nous présentons une étude RMN de dynamique lente de l'ADN, à l'échelle de la microseconde milliseconde. Très peu de travaux ont été fait dans ce domaine en raison de la complexité des techniques mises en œuvre, que nous avons détaillées dans l'introduction générale. De nos jours, un seul type de transition (entre un état principal et un état excité d'énergie bien supérieure) à cette échelle microseconde – milliseconde a été identifiée par RMN dans les ADN-B par l'équipe du Pr Al-Hashimi.

Lors de premières expériences RMN, l'observation de toutes les adénines incluses dans les quatre dodécamères fait apparaître des phénomènes d'élargissement des résonances des protons H2 très particuliers, se produisant dans un contexte de séquence TTAAA.TTTAA. Ces élargissements sont connus pour refléter des mouvements lents (Kennedy et al. 1993) (McAteer et Kennedy 2000). C'est pourquoi nous avons ensuite réalisé des expériences de dispersion de relaxation du carbone 13, qui ont permis de confirmer l'existence d'un l'échange conformationnel lent et d'en préciser certaines caractéristiques.

Dans ce chapitre nous allons tout d'abord exposer la méthodologie que nous avons mise au point pour étudier un dodécamère marqué ^{13}C / ^{15}N . Nous présenterons ensuite les résultats. Nous verrons que nous avons pu identifier un type de mouvement nouveau, dont les caractéristiques diffèrent de celles des des mouvements observés par l'équipe du Pr Al-Hashimi. La caractérisation complète de ce nouveau type de transition nécessiterait des expériences complémentaires. Toutefois, les travaux réalisés dans le cadre de ma thèse constituent un socle très utile pour les études à venir puisqu'ils nous ont permis de maîtriser les différents aspects techniques des expériences de dispersion de relaxation des ADN-B.

2 Méthodologie

2.1 Préparation de l'échantillon

Les oligomères, marqués uniformément (sur tous les carbones et azotes) en $^{13}\text{C}/^{15}\text{N}$ ont été préparés par la société Silantes/Eurisotop (Allemagne) et les oligonucléotides non marqués ont été synthétisés par Eurogentec (Belgique). Les deux brins de séquence 5'-CCGCTTAAACGC-3' et 5'-GCGTTTAAGCGG-3' ont été solubilisés dans un tampon phosphate de sodium à une force ionique de 0.1 (mol/L) et contenant 0.1mM d'EDTA avec un pH 6.5. Les duplexes d'ADN sont préparés en mélangeant les deux brins complémentaires avec un ratio 1:1 dans 450 μL de H_2O . Ensuite les échantillons sont lyophilisés trois fois dans 99.9 % D_2O . Les concentrations finales des oligomères marqués et non marqués sont de 0.85mM et 1.2mM respectivement dans des volumes 180 μL et 500 μL .

2.2 Expériences RMN

2.2.1 Mesure de température de fusion

Une série d'expériences uni-dimensionnelle (1D) avec pré-saturation sur l'eau (H_2O) (zgpr) ont été réalisées à des températures de 5 à 75°C, afin de mesurer la température de fusion. Les expériences ont été effectuées sur les oligomères non marqués.

2.2.2 Mesure des largeurs des raies des protons H2

Afin de discriminer facilement les protons H2 des adénines des protons H6/8 des bases aromatiques qui résonnent dans la même gamme de déplacements chimiques, une série d'expériences 1D T_1 inversion retour (T_{1ir1d}) a été réalisée avec différents délais de relaxation (300ms à 850ms en fonction de la température) entre l'impulsion 180° (inversion) et l'impulsion 90° de lecture avant l'acquisition. Les très importantes différences de temps de relaxation T_1 entre les protons H2 et les

protons H6/H8 permettent d'obtenir des spectres dans lesquels les résonances H6/H8 ont déjà largement relaxées (H6 et H8) tandis que les résonances H2 ne l'ont pas encore fait. Rappelons que dans l'ADN B "canonique", les protons H6 et H8 se trouvent à des distances proches de plusieurs protons des désoxyriboses (H1', H2', H2'', H3' etc ...) alors que le proton H2 est situé beaucoup plus loin (plus de 4 Å) de ces mêmes protons désoxyriboses, de ce fait la relaxation par les mécanismes de type dipôle-dipôle, l'un des deux modes prédominants dans les macromolécules biologiques, est peu efficace pour ces protons H2

Les expériences 1D T₁ρ 1D visant à mesurer les largeurs de raies ont été réalisées aux températures suivantes: 5°, 8°, 10°, 12°, 15°, 18°, 20°, 25°, 28°, 30°, 35°, 40° et 45° C.

2.2.3 HSQC ("*Hétéronuclear Single-Quantum Correlation*") à temps constant

Les déplacements chimiques des protons et carbones ont déjà été déterminés lors de nos précédentes études, (Xu et al. 2014) et lors de la mesure des RDC (Chapitre 3). Ces expériences précédentes avaient été réalisées sur les oligomères en abondance naturelle.

Cependant, afin de confirmer ces déplacements sur le nouvel échantillon marqué, des expériences 2D HSQC à temps constant (Vuister et Bax 1992) ont été réalisées sur l'oligomère enrichi uniformément ¹³C/¹⁵N. Des expériences HSQC optimisées pour différentes régions (régions des bases aromatiques ou des désoxyriboses) ont été enregistrées séparément. Les fenêtres spectrales étaient de 5000 Hz pour la dimension proton et de 5030 Hz dans la dimension carbone. 512 expériences de 2048 points complexes ont été enregistrées avec une accumulation de 16 scans à chaque fois. La séquence d'impulsions utilise des périodes de temps constant ("*constant-time*") pour refocaliser les couplages carbones-carbone (présents dans les cytosines et les thymine pour les bases et dans tous les nucléotides pour les désoxyriboses). Des délais des temps constant de 13.2 ms et 24 ms ont été utilisés pour les régions aromatiques et les sucres respectivement.

2.2.4 Calibration de la force de champ de verrouillage de spin ("*spin lock*") (ω_1) pour les expériences de dispersion de relaxation de type R1ρ

La séquence d'impulsions R1ρ 1D sélective a été décrite dans l'introduction. Ainsi qu'il a été expliqué un champ de radiofréquence de type "*spin lock*" ¹³C est appliqué durant le délai de

relaxation. Il est important de connaître précisément la force de champ correspondante, notamment pour permettre l'extraction des paramètres associés à l'échange. Nous avons utilisé des méthodes de calibration décrites par Arthur Palmer et collaborateurs (Palmer, Kroenke, et Patrick Loria 2001). Dans cette méthode, une série d'expériences de type HSQC temps constant (nous avons surtout travaillé sur la région des bases) est réalisée avec un découplage à onde continue sur le canal carbone pendant le temps d'acquisition. La puissance de découplage est variée en utilisant des atténuations allant de 20db à -4db avec un incrément de 2db. La force de champ (ω_1) peut être déterminée à partir de la relation (34). J_0 étant le couplage scalaire $^1J \text{ } ^1\text{H}-^{13}\text{C}$, J_r étant le couplage apparent qui est relié à la déviation de la position par rapport à la résonance ("Off-resonance") d'une valeur donnée, Ω étant l'écart en Hertz entre la résonance du noyau et la porteuse de l'impulsion.

$$\tan(\theta) = \frac{\omega_1}{\Omega} = \sqrt{\left(\left(\frac{J_0}{J_r}\right)^2 - 1\right)} \quad (34)$$

2.2.5 Calcul de la contribution du Hartmann-Hahn à la relaxation

Comme il a été mentionné dans l'introduction l'un des problèmes associé à l'utilisation d'un champ "spin lock" est la possibilité de transfert de type Hartmann-Hahn (à travers les couplages scalaires) durant cette période. Il est donc nécessaire de calculer les conditions dans lesquelles de tels transferts peuvent se produire.

L'efficacité maximale du transfert Hartmann-Hahn entre deux spins I et S est donnée par l'équation (29) (Bax et Davis 1985) (voir la partie 6.3.2 de l'introduction générale). Selon cette équation le transfert peut être très important (total dans certaines conditions) lorsque les deux spins impliqués ont des résonances proches et sont reliés par une constante de couplage scalaire importante.

Pour calculer l'efficacité du transfert Hartmann-Hahn pour les différents spins ^{13}C des bases adénines, il est nécessaire de connaître les déplacements chimiques des différents carbones impliqués.

Nous avons donc réalisé une expérience 3D TROSY *relayed* HCCH-COSY (Simon, Zanier, et Sattler 2001) (Figure 29) pour mesurer les déplacements chimiques des carbones quaternaires des adénines.

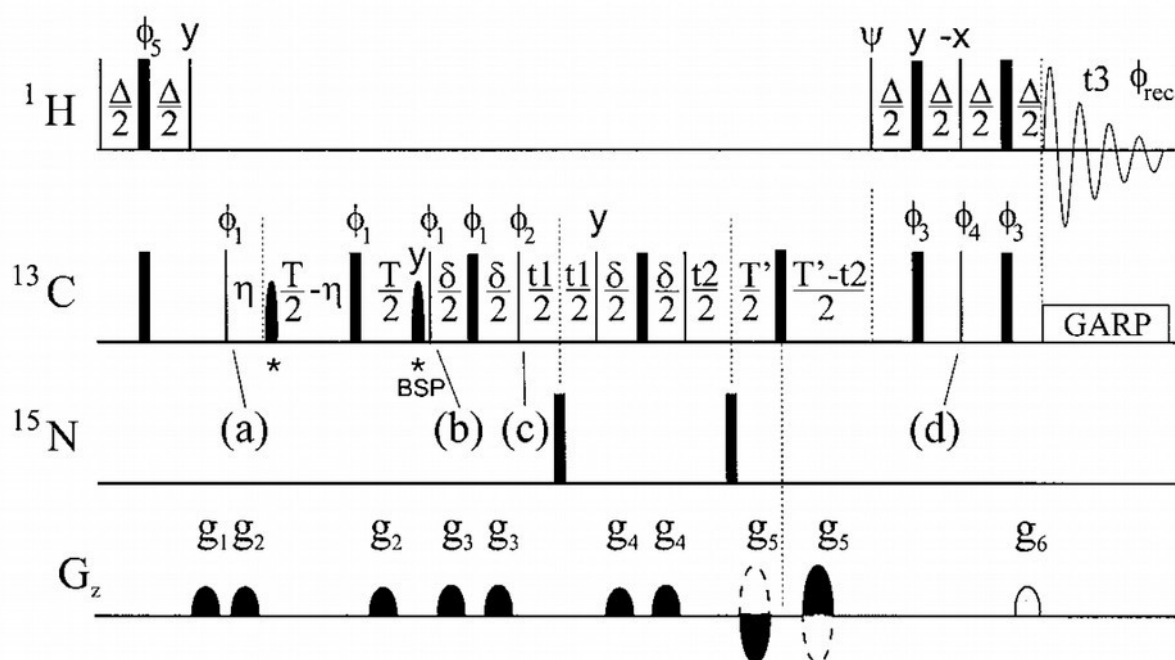


Figure 29 : Séquence d'impulsions de l'expérience 3D TROSY *relayed* HCCH-COSY. Les rectangles minces et épais représentent les impulsions 90° et 180° respectivement. Les impulsions sélectives de type *Band-selective E-BURP-2* (Geen et Freeman 1991) sont indiqués par le symbole (*). Le premier BURP-2 d'une durée de 1.5 ms est appliqué à une fréquence de 100 ppm (*ie*, à 5000 Hz de la porteuse qui est à 140 ppm) excite sélectivement les carbones quaternaires C5 qui résonnent dans cette gamme de déplacements chimiques. L'application du gradient g_2 simultanément mène à la suppression des pics de corrélation H6-C6 des pyrimidines. Le deuxième E-BURB (BSP) pour (Bloch-Siergert Phase) est appliqué pour compenser les déplacements de phase (Sattler, Schleucher, et Griesinger 1999).

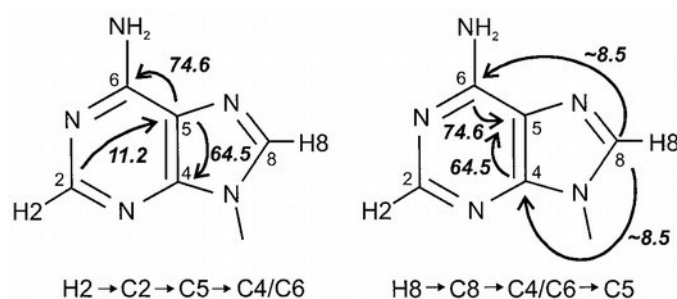


Figure 30 : Représentation schématique du transfert de l'aimantation dans l'expérience TROSY *relayed* HCCH-COSY provenant du H2 (à droite) et du H8 (à gauche). Les chiffres en gras indiquent les constantes de couplages scalaires en Hertz entre les carbones, reliés par les flèches, qui sont à l'origine du transfert de cohérence.

Dans la Figure 29 la séquence d'impulsions de l'expérience est décrite et les transferts d'aimantation impliqués sont indiqués dans la Figure 30.

Pendant l'expérience 3D TROSY *relayed* HCCH-COSY, l'aimantation est transférée simultanément sur les trois carbones aromatiques C4, C5 et C6 (Figure 30) permettant ainsi l'identification de leurs déplacements chimiques. A la position (a) de la séquence d'impulsion (Figure 29), l'aimantation proton carbone en anti-phase est créée pour les couples H8/C8 et H2/C2. Ensuite, les carbones C2 et C8 sont corrélés via les couplages "*long-range*" avec les carbones quaternaires C4, C5 et C6 selon le schéma de la Figure 30. L'impulsion sélective E-BURP appliquée à la fréquence des C5 permet convertir la cohérence en anti-phase des C5-C6 des pyrimidines en une cohérence "*multiple-quantum*". Ces dernières sont supprimées par l'application du gradient g_2 (Figure 29) : ce filtrage mène donc à une suppression des pics de corrélation H6-C6 des pyrimidines. Au point (b) de la Figure 29, des cohérences zéro et double-quantum entre les C8/C4, C8/C6 et C2/C5 sont créées, suivie d'un transfert COSY durant les délais δ via $^1J_{CC}$ entre les noyaux C4, C5 et C6. Durant le t_1 (point c), les déplacements chimiques des tous les carbones des adénines sont enregistrés puis l'aimantation est "renvoyée" aux protons H8 et H2. Les opérateurs produits au point (a), (b) et (c) sont donnés par les équations (36) et (37) pour les carbones C8 et C2 respectivement.

$$\begin{aligned}
 &C8 \rightarrow C4/C6 \rightarrow C5 \rightarrow C4/C6 \rightarrow C8 \\
 &C8_y \rightarrow C8_x \rightarrow +C8_y (36a) \\
 &C8_y \rightarrow -2C8_z C4/6_y \rightarrow -C8_y (36b) \\
 &C8_y \rightarrow -4C8_z C4/6_z C5_x \rightarrow +C8_y (36c)
 \end{aligned} \tag{36}$$

$$\begin{aligned}
 &C2 \rightarrow C5 \rightarrow C4/C6 \rightarrow C5 \rightarrow C2 \\
 &C2_y \rightarrow C2_x \rightarrow +C2_y (37a) \\
 &C2_y \rightarrow -2C2_z C5_y \rightarrow -C2_y (37b) \\
 &C2_y \rightarrow -4C2_z C5_z C4/6_x \rightarrow +C2_y (37c)
 \end{aligned} \tag{37}$$

Les opérateurs indiqués par les équations (36) et (37) montrent que les pics croisés (36b et 37b) sont de signes opposés aux pics diagonaux (36a et 37a) et aux pics reliés croisés (36c et 37c). Une acquisition en 3D est impérative pour que ces pics ne s'annulent pas. Aussi, ces équations montrent que les aimantations croisées des spins sont alignées selon des axes différents : les C6 et C4 selon l'axe z (36c) et les C8 et C5 selon le plan x-y. Le résultat de cette différence d'alignement est que les pics C8 et C5 seront de signes opposés par rapport aux pics C6 et C4. De même, les pics (C2, C4 et C6) d'une part et C5 d'autre part, seront de signes opposés.

L'expérience 3D TROSY *relayed* HCCH-COSY est réalisée avec l'acquisition de 230, 100 et 1024 points complexes dans les trois dimensions, t_1 , t_2 et t_3 respectivement, avec une accumulation de 2 scans à chaque point. Les gammes spectrales étaient de 12577Hz, 7546Hz pour les deux

dimensions carbone t_1 et t_2 respectivement, et de 10000Hz pour la dimension proton t_3 .

2.2.6 Expériences 1D "On resonance" R1 ρ

Des expériences à une dimension sélective du type R1 ρ (Hansen et al. 2009) ont été réalisées pour identifier les spins impliqués dans des processus d'échange conformationnel. La séquence d'impulsions est celle de la Figure 21. Cette expérience est initiée par un transfert de polarisation croisé sélectif de type Hartmann-Hahn hétéronucléaire (ω_{cp}) de ~ 100 Hz (Chiarparin, Pelupessy, et Bodenhausen 1998). Le découplage onde continue (*cw*, "continuous wave") pendant le temps de relaxation est appliqué avec une force de champ de 12 KHz sur le canal ^1H . Ce découplage permet d'éviter l'évolution sous les couplages J_{CH} durant la période de relaxation et permet l'élimination des corrélations croisées dipôles-dipôles(DD)/Anisotropie de déplacement chimique (CSA). Pendant l'acquisition du signal, un découplage de 3.5 KHz de séquence GARP et un découplage de 2.5KHz de séquence WALTZ sont appliqués sur les canaux ^{13}C et ^{15}N respectivement.

10 délais de relaxation ont été utilisés pour chaque expérience { 0, 4, 8, (12 deux fois), 16, 20, 24, (32 deux fois) ms} avec une accumulation de 64 scans pour les carbones C2 des adénines et 32 scans pour les carbones C6/8 de toutes les bases aromatiques. Des valeurs de "spin lock" (ω_{13C}) de 180 – 3600 Hz sont appliquées pendant ces temps de relaxation.

2.3 Spectromètres

Les expériences zgpr, T1ir1d et TROSY *relayed* HCCH-COSY ont été réalisées à Cachan sur un spectromètre AVANCE Bruker 500MHz avec une sonde chaude 5 mm (BBI) . Les expériences HSQC et R1 ρ ont été réalisées sur un spectromètre AVANCE 600MHz avec une cryo-sonde TCI à Gif sur Yvette.

2.4 Traitement et analyse des données

Les expériences zgpr et T1ir1d ont été traitées avec Topspin (Bruker) ; les largeurs à mi-hauteur sont obtenues avec la commande "hwcal" du logiciel.

Les expériences HSQC et HCCH-COSY-relayed TROSY sont traitées avec Topspin (Bruker) et Nmrpipe (Delaglio et al. 1995) ; les spectres ont été analysés et attribués avec le logiciel Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

Les spectres 1D R1ρ ont été traités avec Nmrpipe et les rapports signal/bruit mesurés avec Nmrpipe en utilisant la méthode *DOSY-style 1D-Series (Pseudo-2D)*. Pour chaque expérience, une série d'intensités relatives à l'intensité au temps de relaxation 0 (I_T/I_0) (I_0 : intensité à temps 0 et I_T : intensité à temps T) est ainsi obtenue. L'ensemble de la procédure a été automatisé avec des scripts Python.

Le calcul de la vitesse de relaxation R1ρ (équation (38)) pour chaque résidu est réalisé avec des simulations de type Monte-Carlo.

$$\frac{I_T}{I_0} = e^{(-R_1\rho)T} \quad (38)$$

Celles-ci consistent à générer des jeux de données synthétiques avec une distribution normale de moyenne égale à l'intensité à un temps T et d'une variance $\sigma = \Delta I / 2.354$ où ΔI est le plus grand écart entre les intensités des points dont les mesures expérimentales ont été répétées deux fois (12 et 32ms).

La minimisation des χ^2 est effectuée, suivant l'équation (39) :

$$\chi^2 = \sum_{i=0}^N \left(\frac{Y_i - y_i}{dy_i} \right)^2 \quad (39)$$

Y_i sont les points théoriques donnés par les minimisations utilisant l'algorithme Gauss-Newton pour chaque jeu de données. (y_i) et (dy_i) représentent les points expérimentaux et les erreurs de mesure respectivement.

Toutes les fonctions de minimisation utilisées sont écrites avec le programme **R** (R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria).

La dernière étape est l'ajustement de la courbe des R1ρ en fonction des forces de champ ω_1 . L'expression générale du R1ρ, valable pour tous les régimes d'échange asymétrique ($p_a \gg p_b$) issue de l'équation de Bloch-McConnell (McConnell 1958) est :

$$R_1\rho(\omega_1) = R_1 \cos^2(\theta) + R_2 \sin^2(\theta) + \sin^2(\theta) \left(\frac{p_A p_B \Delta \omega^2 k_{ex}}{(\Omega_A + \Delta \omega)^2 + k_{ex}^2 + \omega_1^2} \right) \quad (40) \quad (\text{Palmer et Massi 2006})$$

Ω_A correspond à la fréquence de résonance de l'état majoritaire. Dans les expériences On-

resonance l'angle d'impulsion θ est de 90° et $\Omega_A = 0$. L'équation 40 peut donc être simplifiée :

$$R_1 \rho(\omega_1) = R_2 + R_{ex} = R_2 + \left(\frac{p_A p_B \Delta \omega^2 k_{ex}}{\Omega_B^2 + k_{ex}^2 + \omega_1^2} \right) \quad (41)$$

En absence d'expérience *Off-resonance* qui permette de résoudre la fréquence Ω_b et $\Delta\omega$, le fit selon un modèle d'échange rapide selon l'équation 42 (Hansen et al. 2009) (Nikolova et al. 2011) (Nikolova, Gottardo, et Al-Hashimi 2012) a été utilisé pour ajuster les courbes des expériences $R_1\rho$ On-resonance et déterminer le k_{ex} et le R_2 .

$$R_1 \rho(\omega_1) = R_2 + R_{ex} = R_2 + \left(\frac{\Phi_{ex} k_{ex}}{k_{ex}^2 + \omega_1^2} \right) \quad (42) \quad \text{où} \quad \text{et} \quad \Phi_{ex} = p_A p_B \Delta \omega^2$$

Ce modèle a permis d'obtenir le k_{ex} , R_2 mais les différents termes contenus dans Φ_{ex} sont confondus et ne peuvent être dissociés.

Une minimisation χ^2 est réalisée en utilisant l'algorithme de Broyden-Fletcher-Goldfarb-Shanno (BFGS) (*Numerical Optimization* 2006). Toutes les équations sont écrites en Python en utilisant l'algorithme (BFGS) implémenté dans le module *Scipy*. Ce modèle a permis d'obtenir les paramètres de la constante d'échange (k_{ex}) et vitesse de relaxation transversale R_2 des noyaux en échanges conformationnel.

3 Résultats et discussion

Pour faciliter la lecture, nous donnerons ici la séquence et la numérotation du dodécamère étudié, appelé Oligo 3:

5'-C₁ C₂ G₃ C₄ T₅ T₆ A₇ A₈ A₉ C₁₀ G₁₁ C₁₂-3'

3'-G₂₄ G₂₃ C₂₂ G₂₁ A₂₀ A₁₉ T₁₈ T₁₇ T₁₆ G₁₅ C₁₄ G₁₃-5'

3.1 Mesure de la largeur à mi-hauteur des résonances des protons H2 des adénines

Les protons H2 des adénines ont un temps de relaxation longitudinale (T_1) long par rapport aux autres protons aromatiques (H6 et H8). Ce phénomène de relaxation T_1 lent des protons H2 adénines est lié à de faibles contributions à la relaxation des mécanismes du type dipôle-dipôle du fait de l'isolement de ces protons relativement aux protons H6 et H8. Par exemple, ces derniers relaxent très efficacement en T_1 du fait de la présence dans leur voisinage des protons H1', H2', H2'' etc.. de leur propre résidu et des résidus adjacents. Ce phénomène de relaxation lente des protons H2 est plus important dans les spectres obtenus dans D₂O que dans ceux obtenus dans H₂O du fait du remplacement des protons échangeables (les plus proches des H2) par des deutérons dans les expériences dans D₂O. En choisissant des délais de relaxation adéquats, il est donc possible d'obtenir des conditions dans lesquelles les protons H6/H8 ont déjà significativement relaxé tandis que les protons H2 n'ont que très peu relaxé. Ainsi il devient facile d'observer sélectivement des résonances H2 des adénines alors que les autres protons des bases sont annulés ou de signes opposés. Cette expérience nous permet donc de pouvoir mesurer les caractéristiques associées à ces résonances sans se soucier des problèmes de superposition qui sont notables dans les spectres 1D standards.

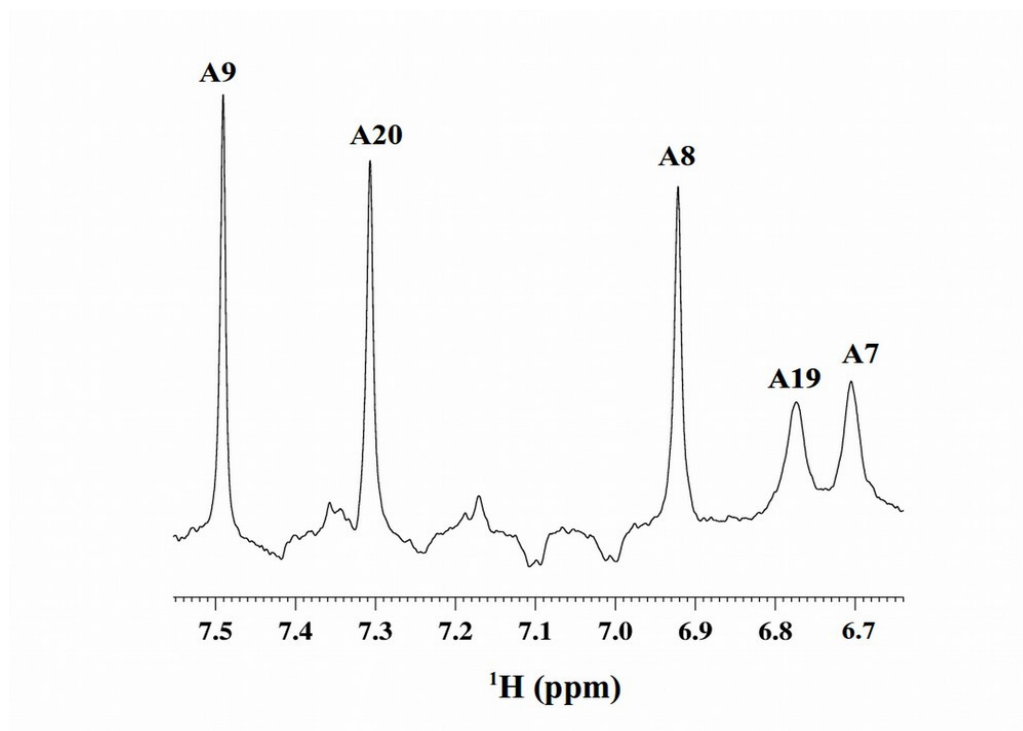


Figure 31 : Région des bases aromatiques du spectre 1D T₁ir1d (500MHz) de l'oligo 3 non enrichi à 25°C. Les raies attribuées correspondent aux protons H2 des adénines.

De cette façon l'expérience T₁ir1d (Figure 31) nous a permis de discriminer facilement les protons H2 des adénines des protons H6 et H8 des pyrimidines ou des purines qui résonnent dans la même région. Ainsi les protons H2 des adénines A9 et A20, qui dans un spectre 1D classique (zgpr) sont complètement superposés avec d'autres résonances des bases aromatiques, deviennent parfaitement observables. La Figure 31 montre la large dispersion des déplacements (δH) chimiques des H2 des adénines. Nous observons aussi des différences importantes dans les largeurs des différents pics. La mesure de variabilité des déplacements chimiques lors de nos précédentes études (Tableau 5) montre que les H2 des adénines ont la plus grande gamme de déplacements chimiques de tous les protons non échangeables de chacun des 4 dodécamères étudiés dans le cadre de cette thèse.

Bases	Proton	Oligo 1	Oligo 2	Oligo 3	Oligo 4
Adénine	H2	0.75	0.45	0.80	0.38
	H8	0.44	0.22	0.21	0.03
Cytosine	H5	0.61	0.39	0.57	0.28
	H6	0.43	0.35	0.32	0.18
Guanines	H8	0.44	0.38	0.34	0.17
Thymines	H8	0.24	0.33	0.20	0.10

Tableau 5 : variabilité des déplacements chimiques en ppm des différents protons non-échangeables des bases aromatiques des quatre dodécamères à 25°C.

Des calculs théoriques de déplacements chimiques (Giessner-Prettre et Pullman 1970) ont montré que la contribution du courant de cycle des bases aromatiques est plus importante pour les protons H2 des adénines (1.33ppm) que pour les H8 des guanines (0.73ppm), ou les H6 des cytosines (0.32ppm) et des thymines (0.13ppm), d'où la plus grande distribution des déplacements chimiques des protons H2 comparé aux autres protons non-échangeables des bases aromatiques. Cette sensibilité des protons H2 des adénines aux courants de cycle, traduite par la variabilité de δH , confère à ces protons un bon potentiel prédictif de l'échange conformationnel via l'étude des élargissements des résonances de ces noyaux (McAteer et Kennedy 2000)

Les adénines qui montrent des élargissements importants dans les spectres de la Figure 31 sont toutes deux incluses dans des enchaînements dinucléotidiques de type TpA et sont les adénines A7 et A19. On peut remarquer que par rapport aux autres adénines, leurs largeurs à mi-hauteur sont beaucoup plus importantes.

La sensibilité du proton H2 au courant de cycle des bases adjacentes pourrait être avancée pour expliquer les différences des largeurs de résonances. En fait, selon la simulation de l'équation de McConnell (Palmer, Kroenke, et Patrick Loria 2001) que nous avons détaillée dans l'introduction, les élargissements des résonances sont régis par deux constantes (k_{ex} et $\Delta\omega$). L'hypothèse que les constantes d'échange k_{ex} soient identiques pour toutes les H2 des adénines et que c'est le $\Delta\omega = \omega_A - \omega_B$ qui est modulé par le courant de cycle qui accroît les largeurs de raie n'est pas recevable, puisque les raies des H2 des adénines situées entre deux bases puriques dont l'effet du courant de cycle est plus important (A8 dans la séquence AAA et A20 dans la séquence AAG) sont plus

finer que celles des H2 des pas TpA. En outre, des travaux antérieurs (McAteer et Kennedy 2000) qui visaient à quantifier les élargissements de résonances des pas TpA dans les 16 contextes tétranucléotidiques possibles, ont montré que les élargissements de résonances des H2 des adénines des pas TpA ne peuvent pas s'expliquer par l'effet du courant de cycle. Par exemple, ces auteurs ont montré que les largeurs de raies H2 des adénines des pas TpA dans les contextes TTAG et TTAC sont identiques et deux fois moins importantes que dans le contexte TTAA mais que les largeurs des raies H2 dans ATAA et TTAT sont comparables.

Les phénomènes d'élargissement des résonances H2 des adénines des TpA et plus particulièrement dans le contexte TTAA sont connus depuis 1985 (J.-F. Lefevre, Lane, et Jardetzky 1985). L'hypothèse de l'existence d'un échange conformationnel a été avancée pour expliquer cet élargissement. L'enchaînement TpA est en effet souvent considéré comme flexible, les interactions d'empilement étant particulièrement faibles. Le groupe de Reid publia en 1993 (Kennedy et al. 1993) un article où les différents mécanismes (relaxation de type paramagnétique privilégiée, interactions dipôle-dipôle, échange des protons iminos correspondant accéléré, etc ...) qui pourraient expliquer la largeur de raie de l'adénine dans le pas TpA contenu dans un hexadécamère palindromique de séquence $[d(CGAGGTTTAAACCTCG)]_2$, sont passés en revue et discutés. Cette publication a montré de façon convaincante que le phénomène ne peut être expliqué par la contribution de couplages scalaires particuliers proton-proton (ou proton-phosphore) puisque les H2 n'ont aucun voisin de ce type à moins de 5 liaisons chimiques. De même, les interactions dipôle-dipôle ne peuvent expliquer l'élargissement excessif de ce proton, puisque les distances issues des spectres NOESY montrent que le proton le plus proche du H2 élargi se trouve à 4.3 Å. En outre, la mesure de la vitesse de relaxation longitudinale (T_1) des protons H2 en présence de 0.8mM d'EDTA (afin d'éliminer des interactions de l'ADN avec d'éventuels ions paramagnétiques) montre que le proton H2 d'intérêt a le T_1 le plus long, soit 3.8 secondes, comparé aux autres adénines de la séquence dont les T_1 sont compris entre 2.7 et 3.2 secondes. Cette expérience élimine une éventuelle proximité avec un ion paramagnétique qui contribuerait à la relaxation et expliquerait l'élargissement de la résonance. Dans cet article, les auteurs ont conclu que seul un échange conformationnel peut expliquer l'élargissement caractéristique de la résonance de l'adénine du pas TpA dans le contexte TTAAA.

Dans le cadre de notre travail, nous avons suivi les variations des largeurs de raies et de déplacements chimiques des protons H2 de l'Oligo 3 en faisant varier la température. La Figure 32

(à gauche) montre qu'à 5°C, les protons H2 des adénines ont des largeurs de raies de 5Hz à 6Hz à mi-hauteur, puis lorsque la température augmente on observe une très légère diminution de celles-ci. Par contre les résonances A7 et A19 (celles des enchaînements TpA) montrent un comportement très différent avec la température. Les largeurs de raies augmentent avec la température dans un premier temps pour atteindre un maximum à 25°C, puis elles diminuent jusqu'à 45°C pour en fait retrouver leur largeurs initiales

Ce phénomène s'accorde très bien avec l'existence d'un échange conformationnel de ces protons entre deux états A et B. Cet échange conformationnel est quasi inexistant à 5°C, et l'augmentation de la température à 15°C rend la dynamique conformationnelle perceptible sur les élargissements de raies. L'élargissement devient de plus en plus important jusqu'à 25°C. Au-delà de cette température l'élargissement décroît. Il s'agit d'un profil caractéristique d'une variation de la cinétique de l'échange, passant du très lent à 5° à l'intermédiaire puis au rapide, mais notons aussi que l'augmentation de température nous rapproche de la température de fusion de cette séquence qui est de 57°C (Figure 32 à droite). À des températures élevées la structure change et la dynamique spécifique d'un double brin pourrait être prohibée (McAteer et Kennedy 2000).

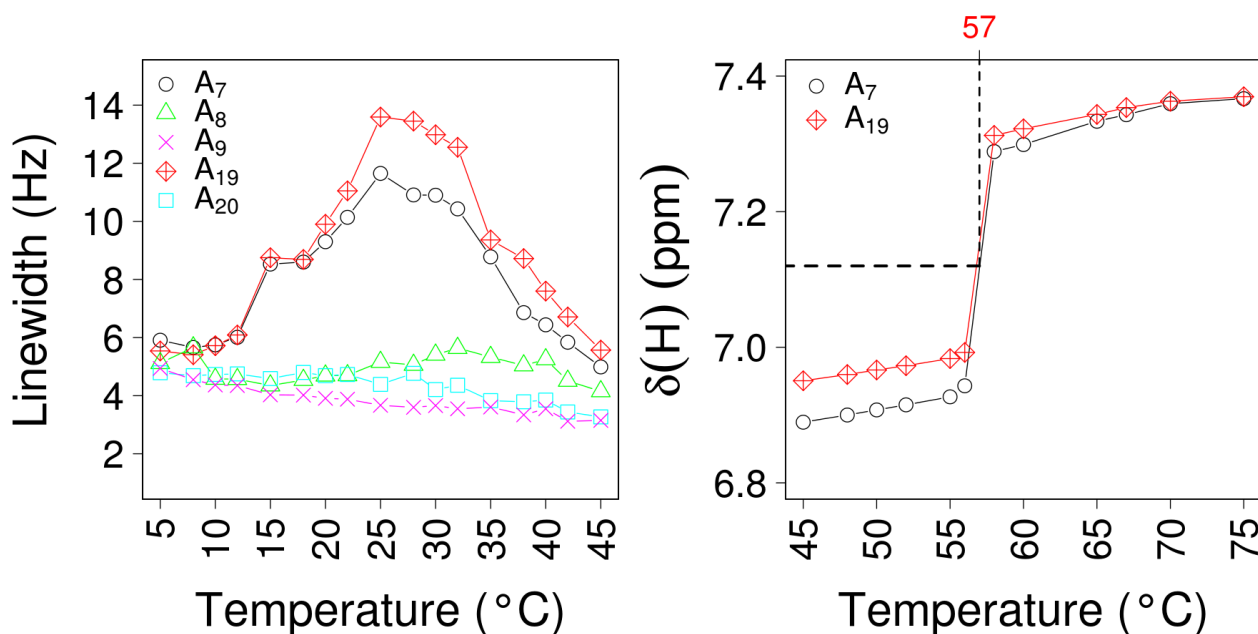


Figure 32: Largeur à mi-hauteur des raies H2 des adénines en fonction de la température (à gauche) et déplacements chimiques (δH_2) à différentes températures (à droite) pour les protons H2 de l'Oligo 3 non enrichi. La légende en haut à gauche montre les différents symboles utilisés pour chaque proton.

Le suivi des largeurs de résonances des protons H2 des adénines, du fait de leur sensibilité à l'environnement chimique, montre donc probablement l'existence d'un mouvement lent. Ces mouvements se situent très probablement dans les échelles de temps de la micro- milliseconde, puisqu'ils sont capables de moduler les largeurs de raies. Les résultats obtenus ici suggèrent que la population de l'état majoritaire p_A est en large excès comparé à l'état minoritaire p_B . En effet seul l'état majoritaire est visible sur les spectres 1D et 2D (échange asymétrique). Dans le but de mieux appréhender les caractéristiques thermodynamiques et physico-chimiques de cet échange conformationnel, nous avons effectué des expériences de dispersion de relaxation. Ces expériences sont un outil puissant dans la mise en évidence des mouvements lents dans les macromolécules et la résolution des structures associées.

3.2 Dispersion de relaxation ^{13}C

Les protons H2 sont portés par les carbones C2 des bases. Nous avons donc mené des expériences pour mesurer les paramètres RMN et notamment les vitesses de relaxation transverse ^{13}C associées aux carbones C2 aromatiques ainsi qu'aux autres carbones aromatiques C8 des purines et C6 des pyrimidines. Ce type de mesure est réalisable sur notre oligomère enrichi uniformément en ^{13}C et ^{15}N (ce dernier noyau a toutefois été peu utilisé dans nos expériences). Nous avons utilisé des expériences de dispersion de relaxation du type R1 ρ en carbone, qui permettent de mettre en évidence les processus d'échange conformationnel et le cas échéant de les quantifier en extrayant des informations d'ordre cinétique, thermodynamique et conformationnel. Les expériences que nous avons réalisées sont les expériences de dispersion de relaxation de type "On-Resonance" R1 ρ sur les différents carbones des bases aromatique. Ces séquences d'impulsions, comme il est expliqué dans l'introduction générale ont été mises au point par le groupe de Lewis Kay sur les groupes N-H des protéines (azote) (Korzhnev, Orekhov, et Kay 2005) puis adaptées pour l'étude des carbones dans les acides nucléiques par le groupe de Hashim Al-Hashimi (Hansen et al. 2009). Notre premier objectif était de mesurer la dispersion de relaxation des carbones C2 des adénines afin de déterminer si le phénomène d'échange mis en évidence sur les protons H2 était également perceptible en observant les carbones C2.

Le premier travail que nous avons effectué a consisté à calibrer les forces de champ "spin lock" ^{13}C

appliqué durant les périodes de relaxation de la séquence "On-Resonance" R1 ρ . Il était aussi important de calculer la contribution éventuelle des phénomènes de type Hartmann-Hahn qui peuvent se produire durant la période de "spin-lock". Ces évaluations nous permettront de bien choisir les forces de champ pour lesquelles cette contribution est négligeable.

3.2.1 Calibration de la force de champ

L'extraction précise des paramètres d'échanges associés aux équilibres conformationnels étudiés nécessitent de calibrer la force de champ de radiofréquence ω_1 utilisée pour le verrouillage de spin ("spin-lock") appliquée sur le carbone ^{13}C durant la période de relaxation (Palmer, Kroenke, and Patrick Loria 2001). Cette valeur est nécessaire pour déterminer les paramètres à partir des équations d'échange conformationnel puisque la vitesse d'échange R_{ex} est modulée par la force de champ appliqué pendant le temps de relaxation (voir Introduction). De plus le calcul des contributions de type Hartmann-Hahn à la relaxation nécessite la connaissance des valeurs exactes de ces forces de champ. La méthode que nous avons utilisée pour cette calibration est la méthode de découplage à onde-continue (cw) "Off-resonance" décrite dans la partie méthodologie. La Figure 33 montre une partie d'un spectre HSQC découplé avec une atténuation de 20 décibel (db).

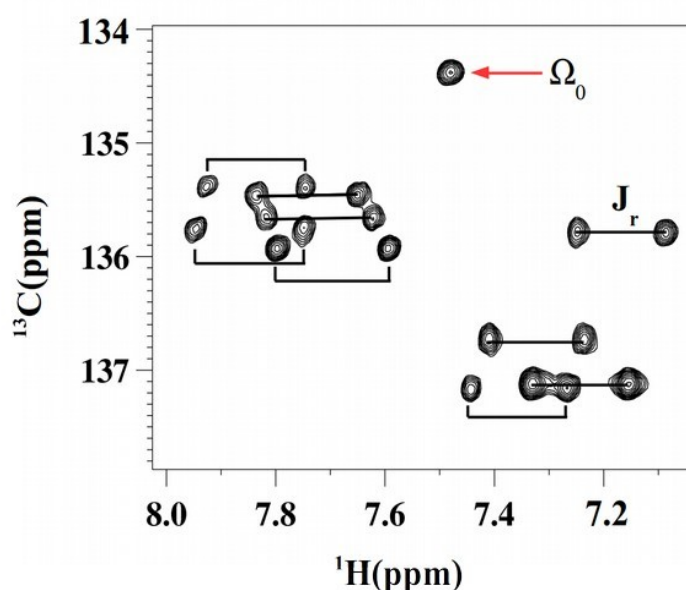


Figure 33: Région des H6/8-C6/8 des bases aromatiques d'un spectre HSQC (en mode temps constant) à 600MHz de l'oligomère enrichi $^{15}\text{N}/^{13}\text{C}$ à 25°C découplé en mode onde continue avec

une atténuation de 20db. Ω_0 représente la fréquence de la porteuse de l'onde de découplage, J_r est le découplage partiel.

Le pic "On-resonance" indiqué et situé au niveau de la porteuse est complètement découplé comme attendu. Par contre les autres pics étant décalés par rapport à la fréquence porteuse ("Off-resonance") dans la dimension carbone, des découplages partiels se produisent et les valeurs de couplages dérivées pour chaque pic sont mesurés. En utilisant l'équation (16), la force de champ est calculée pour chaque valeur de l'atténuation (P) comme dans l'exemple de la Figure 34 à gauche.

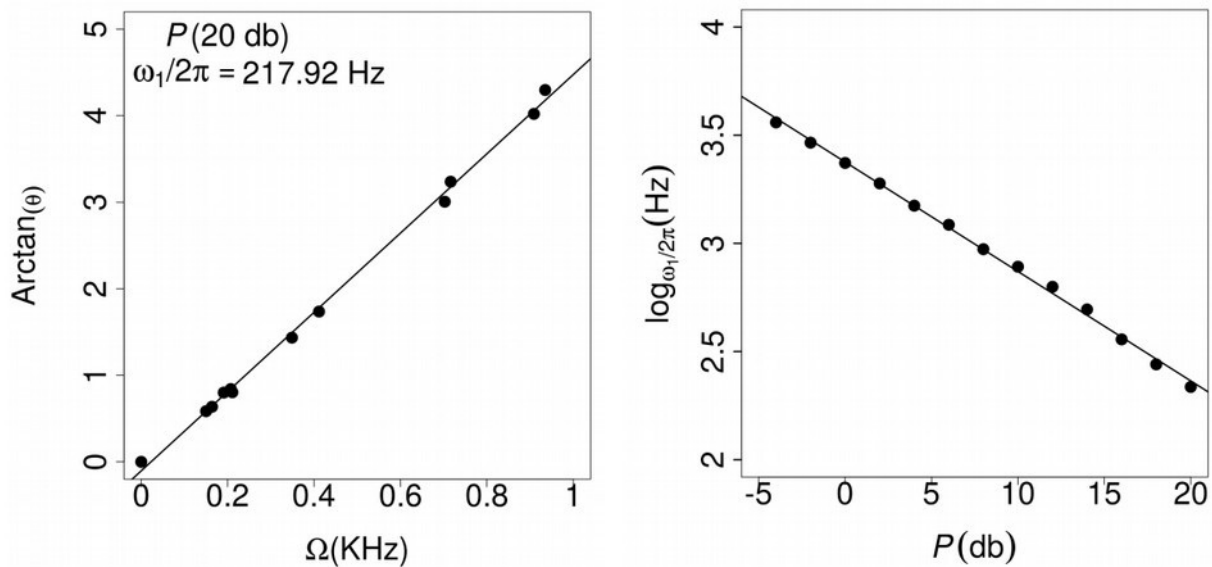


Figure 34 : Graphique à gauche représentant l'arc-tangente de l'angle de l'impulsion (cw) à 20db (0.01 watts) pendant l'acquisition en fonction de l'écart par rapport à la porteuse (Ω). Le graphique à droite représente le logarithme des forces de champ en fonction des atténuations. Les deux droites tracées représentées en ligne continue sont obtenues par minimisation des moindres carrés des points expérimentaux.

Le coefficient de corrélation (R Pearson) des logarithmes de 13 valeurs ($\omega_1/2\pi$) en fonction des atténuations en (db) est de 0.99 (Figure 34 à droite). Avec une corrélation aussi bonne, nous avons pu facilement extraire la fonction (42) qui nous donne la force de champ pour n'importe quelle valeur d'atténuation.

$$\frac{\omega_1}{2\pi} = 10^{-0.05 P_{\text{db}} + 3.38} \quad (42)$$

3.2.2 Calcul de la contribution de Hartmann-Hahn à la relaxation

Il existe une difficulté additionnelle à contourner quand on utilise les séquences de type R1 ρ sur des acides nucléiques uniformément enrichis. Avec les expériences de type R1 ρ , comme avec toutes les expériences où on travaille avec des conditions de verrouillage de spin (*spin lock*), des transferts de type Hartmann-Hahn peuvent se produire, lorsque certaines conditions sont remplies. Dans ce dernier cas les mesures de relaxations sont faussées (Hansen et al. 2009). De tels événements sont perceptibles sur les courbes de décroissance de l'intensité du signal en fonction du délai de relaxation, et causent des oscillations plus ou moins importantes du signal (Mulder et Akke 2003) (Lundström et Akke 2005).

Ainsi, en utilisant des échantillons uniformément marqués tels le dodécamère étudié dans ce travail, une attention toute particulière doit être portée aux forces de champ et aux fréquences utilisées afin d'éviter les conditions de transfert Hartmann-Hahn qui pourraient contribuer à la relaxation en raison des couplages scalaires carbone-carbone (Hansen et al. 2009). L'efficacité des transferts Hartmann-Hahn (A_{HHAH}) des C2/C8 peut être calculée pour l'ensemble des adénines selon l'équation (29). L'application de cette relation implique la connaissance complète des couplages et des déplacements chimiques des systèmes de spin concernés. L'efficacité du transfert A_{HHAH} peut être important notamment dans le cas des C2 des adénines et des C4 du fait de leur proximité de déplacements chimiques (148-157 ppm pour les C2 et 149-154 ppm pour les C4 (Data from Al-Hashimi Lab) (Figure 35), le couplage entre les deux carbones étant toutefois assez faible ($J=1\text{Hz}$). Un autre transfert A_{HHAH} possiblement effectif, toujours pour les adénines, est le transfert entre le C5 et le C2 ($^3J_{\text{C2C5}}$ est de l'ordre de 11Hz), les déplacements chimiques associés étant 118-122 ppm pour les C5.

Nous avons donc déterminé les déplacements chimiques des carbones quaternaires C4, C5, C6 des purines, qui peuvent être obtenus avec des expériences de type 3D TROSY *relayed* HCCH-COSY (Simon, Zanier, et Sattler 2001) (Figure 35)

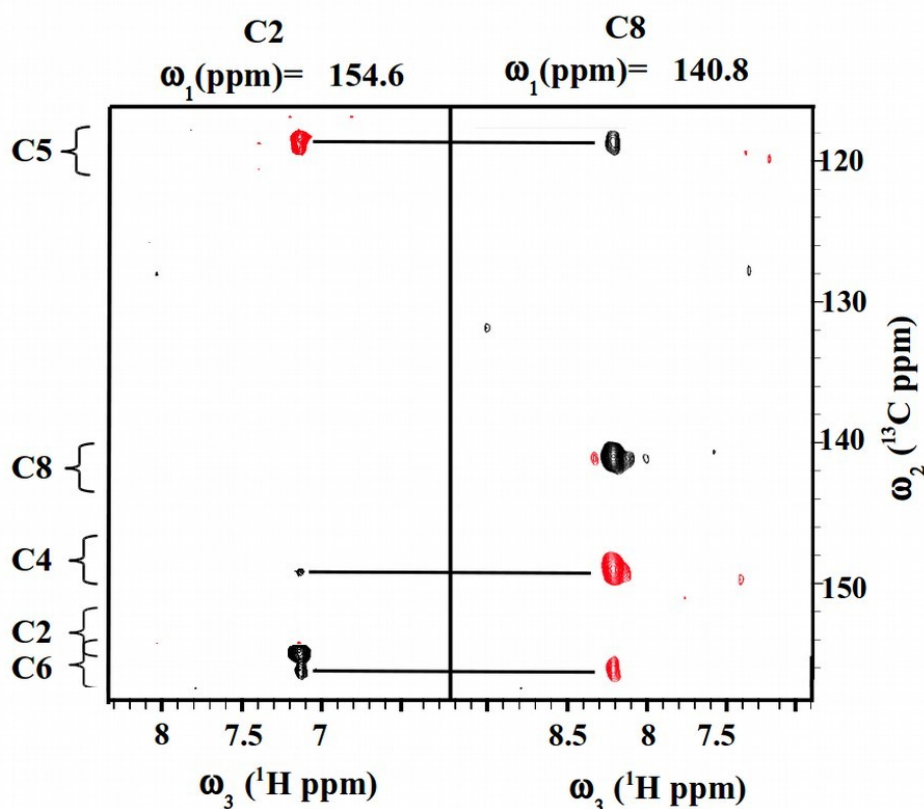


Figure 35: Deux plans du spectre 3D TROSY *relayed* HCCH-COSY à 25°C de l'oligo 3 marqué (500MHz) . Les plans représentent les projections 2D du TROSY *relayed* HCCH-COSY pris aux fréquences des carbones C2 et C8 du même résidu A8. Le panneau de gauche représente le plan du spectres 3D à $\omega_2 = 154.6$ ppm qui est le déplacement chimique du carbone C2 du A8. Les carbones C6, C2 et C4 sont de signe positif (pics en noir) et en rouge le pic C5 est de signe négatif. A droite, la projection 2D à $\omega_2 = 140.8$ ppm correspondant au δC du C8 du même résidu. Les pics C8 et C5 en noir sont de signe positif tandis que les pics C4 et C6 en rouge sont de signe négatif.

L'obtention des déplacements chimiques a permis le calcul des A_{HABA} pour les différents couples de carbones et nous avons pu montrer que cette contribution est toujours inférieure à 0.01 % pour l'ensemble des forces de champ utilisées lors de cette étude (Tableau 6), Cette valeur limite est celle au dessus de laquelle les données ne doivent pas être prises en compte du fait de la présence possible de transferts Hartmann-Hahn (Hansen et al. 2009).

De la même façon, les transferts Hartmann-Hahn A_{HABA} C6-C5 des cytosines peuvent être calculés en utilisant les déplacements chimiques des C5 directement extrait des HSQC. Pour les guanines et les thymines, les déplacements chimiques des carbones quaternaires n'ont pas encore été déterminés mais l'observation de relaxations de type monoexponentielle telle que celles présentées dans la Figure 37 est la preuve que la contribution du A_{HABA} à la relaxation est insignifiante (Hansen et al. 2009). Dans le cas contraire des ondulations seraient visibles sur les données de relaxation (Mulder

et Akke 2003) (Lundström et Akke 2005).

w1(Hz)	C2-C4 $J_{cc} = -1$	C2-C6 $J_{cc} = -1$	C2-C5 $J_{cc} = 11$	C8-C6 $J_{cc} = 8.5$	C8-C5 $J_{cc} = 8.5$
3790	0,001	0,004	0,015	0,000	0,002
2884	0,000	0,005	0,005	0,000	0,001
2290	0,000	0,005	0,001	0,000	0,001
1760	0,003	0,644	0,057	0,000	0,001
1445	0,002	0,003	0,000	0,000	0,000
1148	0,000	0,000	0,004	0,000	0,000
932	0,001	0,001	0,006	0,000	0,000
724	0,001	0,972	0,001	0,000	0,000
601	0,219	0,004	0,995	0,000	0,000
575	0,000	0,001	0,001	0,000	0,000
457	0,001	0,002	0,002	0,000	0,000
363	0,003	0,000	0,001	0,000	0,000
323	0,000	0,000	0,002	0,000	0,000
280	0,098	0,009	0,155	0,000	0,000
257	0,003	0,004	0,002	0,000	0,000
229	0,000	0,002	0,002	0,000	0,000

Tableau 6: Exemple de calcul de la contribution Hartmann-Hahn à la relaxation des C2 et C8 de l'adénine A9. Les valeurs AHAHA sont indiquées pour chaque force de champ (ω_1) et pour chaque couple de carbone. Les calculs ont été effectués avec les couplages scalaires indiqués dans le tableau et les déplacements chimiques mesurés dans les spectres 3D TROSY relayed. Les valeurs et les forces de champ correspondantes pour lesquelles la contribution du A_{HAHA} est significative (en gras), n'ont pas été utilisées dans les expériences R1p.

3.2.3 Dispersion de relaxation "On resonance" R1p

La Figure 36 montre la région des corrélations H2-C2 observées dans un spectre HSQC (temps-constant). Les expériences R1p utilisées pour mesurer les vitesses de relaxation dans le repère tournant (R1p) sont de type 1D et sont sélectives de chaque système de spin ; pour chaque expérience les fréquences carbone et proton impliquées sont utilisées. La sélectivité est assurée par des transferts Hartmann-Hahn hétéronucléaire au début de la séquence avec des excitations aux fréquences exactes du proton et de son carbone associé en utilisant les séquences développées par le groupe de Bodenhausen (Chiarparin, Pelupessy, et Bodenhausen 1998). Le résultat d'une telle excitation est montré sous forme de spectre 1D dans la Figure 36 à l'intérieur de la région du spectre 2D HSQC.

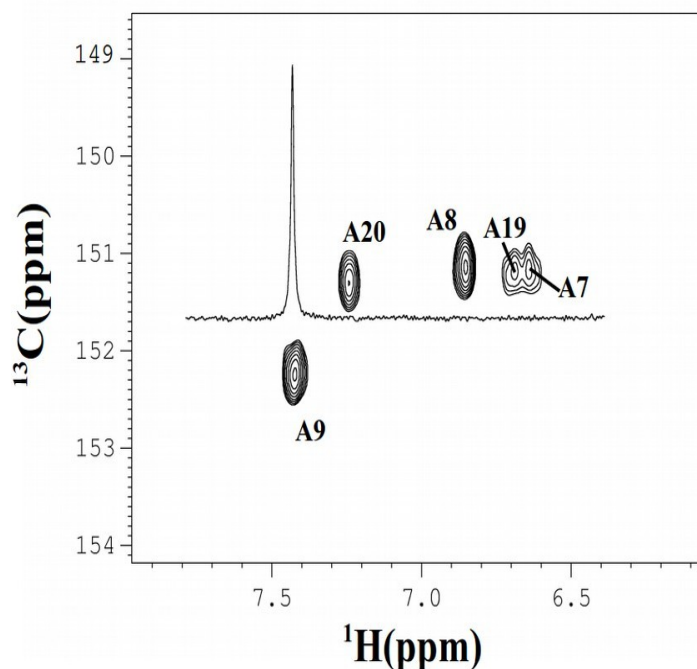


Figure 36 : Région des H2-C2 des adénines d'un spectre ^{13}C - ^1H HSQC (temps-constant) (600MHz) obtenu sur l'échantillon uniformément enrichi à 25°C. Le spectre 1D superposé aux fréquences protons de l'expérience HSQC représente l'expérience sélective R1 ρ .

Les vitesses de relaxation de chaque carbone sont extraites des mesures des intensités des pics obtenues avec différents délais de relaxation (Voir Figure 37). Ces courbes ont été recalculées avec des fonctions de type monoexponentielle. L'intensité de chaque résonance pour chacun des différents délais de relaxation a été normalisée par rapport à l'intensité à temps de relaxation nul. La vitesse est déterminée avec l'équation (38) selon des méthodes de type Monte-Carlo décrites dans Matériels et Méthodes. La Figure 37, montre un exemple de relaxation du carbone C2 obtenue avec différentes valeurs de force du champ "*spin lock*" ^{13}C (ω_1) appliquées pendant les différents temps de relaxation. Les expériences étant sélectives, la relaxation de chaque carbone est réalisée séparément. Une cumulation de 64 scans était nécessaire pour obtenir un rapport signal sur bruit d'environ 30 pour les résidu A7 et A19 et ainsi avoir une bonne précision dans les mesures.

La détermination des vitesses de relaxation à différentes forces de champ "*spin lock*" permet de tracer l'évolution des vitesses de relaxation en fonction de ces valeurs. Comme mentionné dans l'introduction, l'existence de processus d'échange conformationnel dans des fréquences compatibles avec la gamme de forces de champ utilisées se traduit par une variation de la vitesse de relaxation

en fonction de la force du champ "spin lock".

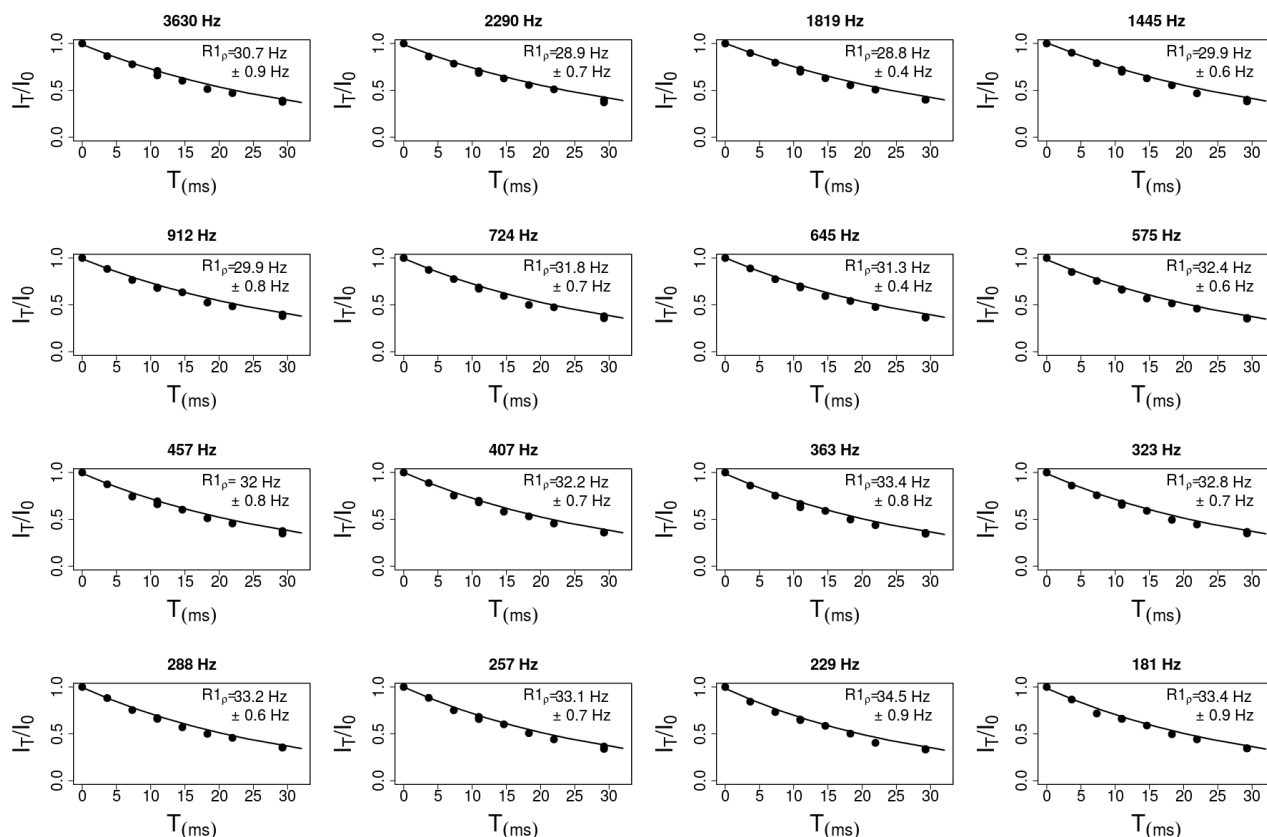


Figure 37 : Exemple des *fits* monexponentiels des expériences $R1_\rho$. Chaque graphique représente l'intensité relative I_T/I_0 en ordonnée en fonction du temps de relaxation pour les différentes forces de champ (titre des graphiques). Les points représentent les mesures expérimentales, et les courbes les *fits* théoriques. La vitesse de relaxation $R1_\rho$ est donnée en légende dans chaque graphique avec l'erreur sur la mesure estimée par les simulations Monte-Carlo. Les données ont été obtenues les adénines A7 de l'oligomère uniformément enrichi en travaillant à 25°C sur le spectromètre 600MHz.

Les résultats sur les différentes vitesses de relaxation des expériences $R1_\rho$ en fonction de la force de champ (Figure 38 & Tableau 7) montrent clairement l'existence de mouvements lents dans les C2 des adénines A7 et A19 avec des vitesses d'échange de 331 et 611 Hz respectivement. Les C2 des adénines A8, A9 et A20 ne montrent pas le profil caractéristique de la présence de mouvements lents puisque l'on n'observe pas de variation des vitesses de relaxation en fonction de la force de champ. Les expériences réalisées ici ("On-resonance") permettent ainsi de distinguer les noyaux en échange conformationnel de ceux qui ne le sont pas (Nikolova et al. 2012). Les erreurs sur les vitesses d'échange et les relaxation transversales (Tableau 7) sont directement extraites des

minimisations Monte-Carlo.

Les données obtenues avec la RMN du carbone confirment donc parfaitement les résultats obtenus en RMN du proton puisque les deux carbones en échange lent sont ceux des résidus TpA dont les protons montraient également des élargissements. Mais, de plus, nous avons pu déterminer pour chaque carbone les paramètres k_{ex} et R_{ex} associés à l'échange conformationnel.

Dans un deuxième temps, des expériences "*off-resonance*" R1 ρ devraient nous permettre de déterminer l'ensemble des paramètres en travaillant à un seul champ magnétique (p_A , p_B et $\Delta(\omega)$) (Nikolova et al. 2012). Néanmoins nous n'avons pas réalisé toutes les expériences nécessaires dans le cadre de ce travail.

Résidu	R_2 (Hz)	Erreur Standard	k_{ex}	Erreur Standard
A7	28.98	0.10	331.78	84.18
A8	30.10	0.06	--	--
A9	29.83	0.06	--	--
A19	28.42	0.05	611.56	62.00
A20	29.56	0.06	--	--

Tableau 7 : paramètres d'échanges conformationnels, des carbones C2 des adénines, R_2 et k_{ex} ainsi que l'erreur standard extraite des minimisations selon l'équation 42. Les résidus en échange conformationnel sont mentionnés en bleu gras. Les résidus qui ne sont pas en échange dans la gamme de temps visible du R1 ρ , sont en noir, la constante d'échange nulle est mentionnée en (- -).

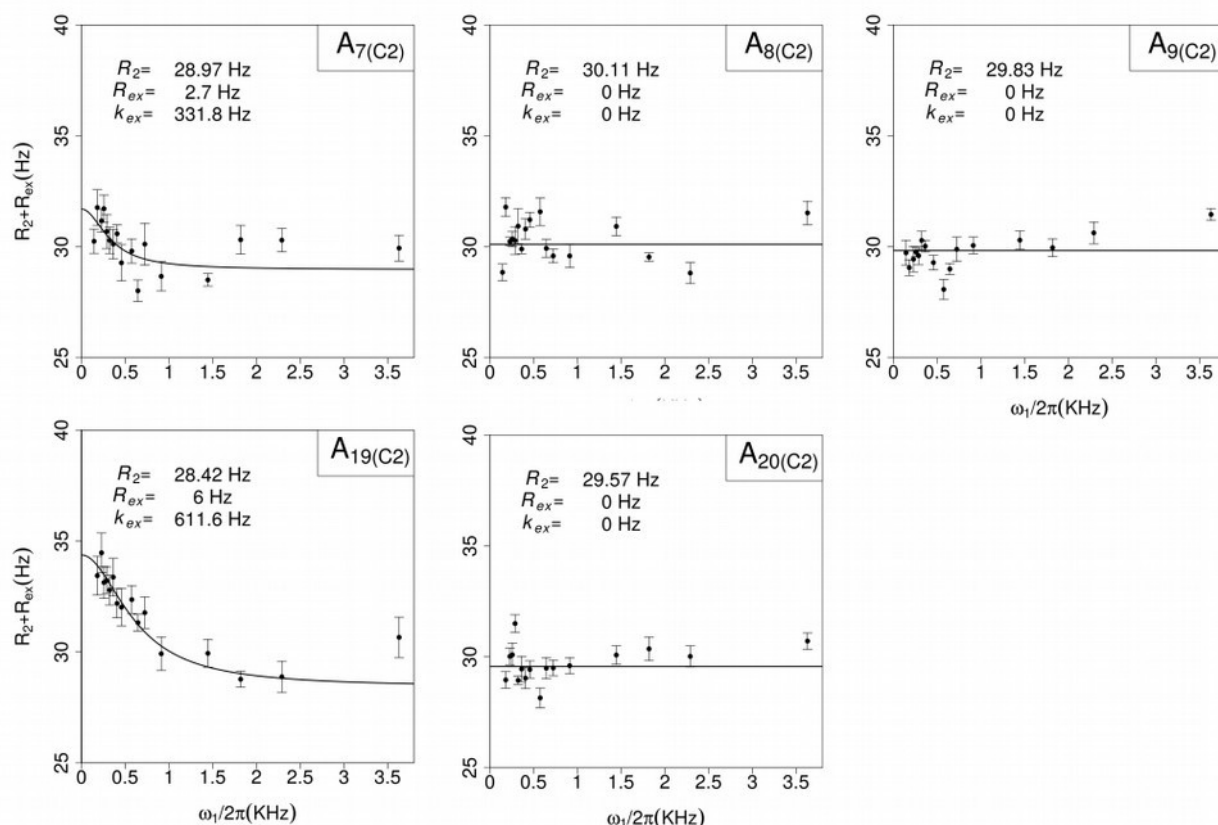


Figure 38 : Vitesses de relaxation $R_1\rho=R_2+R_{ex}$ (en Hertz) en fonction de la force de champ ($\omega_1/2\pi$ en Kilo-Hertz) appliquées sur les différents (C2) des adénines (légende en haut à gauche). La fréquence de relaxation transversale ($R_2=1/T_2$), la constante d'échange (k_{ex}) et la contribution de l'échange à la relaxation (R_{ex}) pour chaque noyau sont mentionnées dans chaque graphique. Les points représentent les mesures expérimentales avec les barres d'erreur correspondantes, les courbes sont donnée par l'équation 42 .

Nous avons également sondé l'existence de mouvements lents sur les carbones C8 des adénines ainsi que sur les C6 des thymine complémentaires aux adénines. Nous avons également recherché l'existence de mouvements sur la cytosine 10 et la guanine 11. Notons que les résonances des carbones C8 des adénines A9 et A20 d'une part et des adénines A7 et A19 d'autre part sont superposées ainsi que les C6 des thymine T16 et T18 (Figure 39).

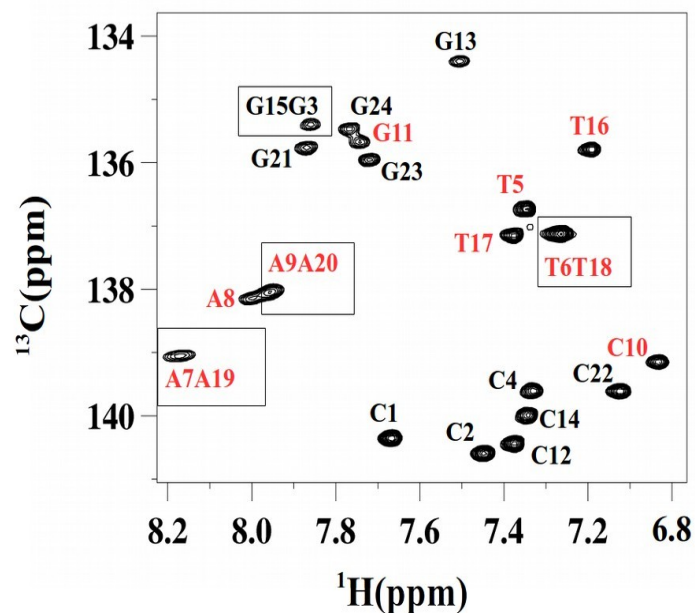


Figure 39: Région des H6/8-C6/8 des bases aromatiques d'un spectre HSQC (temps constant) à 25° de l'oligo 3 enrichi (600MHz). Les attributions en rouge représentent les carbones étudiés en dispersion-relaxation et les pics encadrés représentent les résonances superposées.

Les résultats des expériences R1ρ sur les carbones que nous venons d'énumérer sont présentés dans la Figure 40 et le Tableau 8. Ces données, montrent l'existence de mouvements lents sur le carbone C8 de l'adénine A8. De plus les mesures sur les signaux superposées A7/A19, A9/A20 et T6/T18 montrent la présence d'échanges conformationnels dans la gamme de temps explorée dans au moins une base de chacun des groupes.

Résidu	R ₂ (Hz)	Erreur Standard	k _{ex}	Erreur Standard
T5	30.20	0.07	--	--
T6_T18	31.05	0.10	459.62	44.23
A7_A19	26.47	0.10	573.19	83.21
A8	26.11	0.12	692.6	71.72
A9_A20	26.46	0.10	608.5	56.23
C10	32.18	0.11	821.16	89.65
G11	25.23	0.06	--	--
T16	29.64	0.60	--	--
T17	31.19	0.06	--	--

Tableau 8: paramètres d'échanges conformationnels, des carbone C6/8 des bases aromatique, R₂ et k_{ex} ainsi que l'erreur standard extraite des minimisations selon l'équation 42. Les résidus en échange conformationnel sont mentionnés en bleu gras. Les résidus qui ne sont pas en échange dans la gamme de temps visible du R1ρ, sont en noir, la constante d'échange nulle est mentionnée en (--).

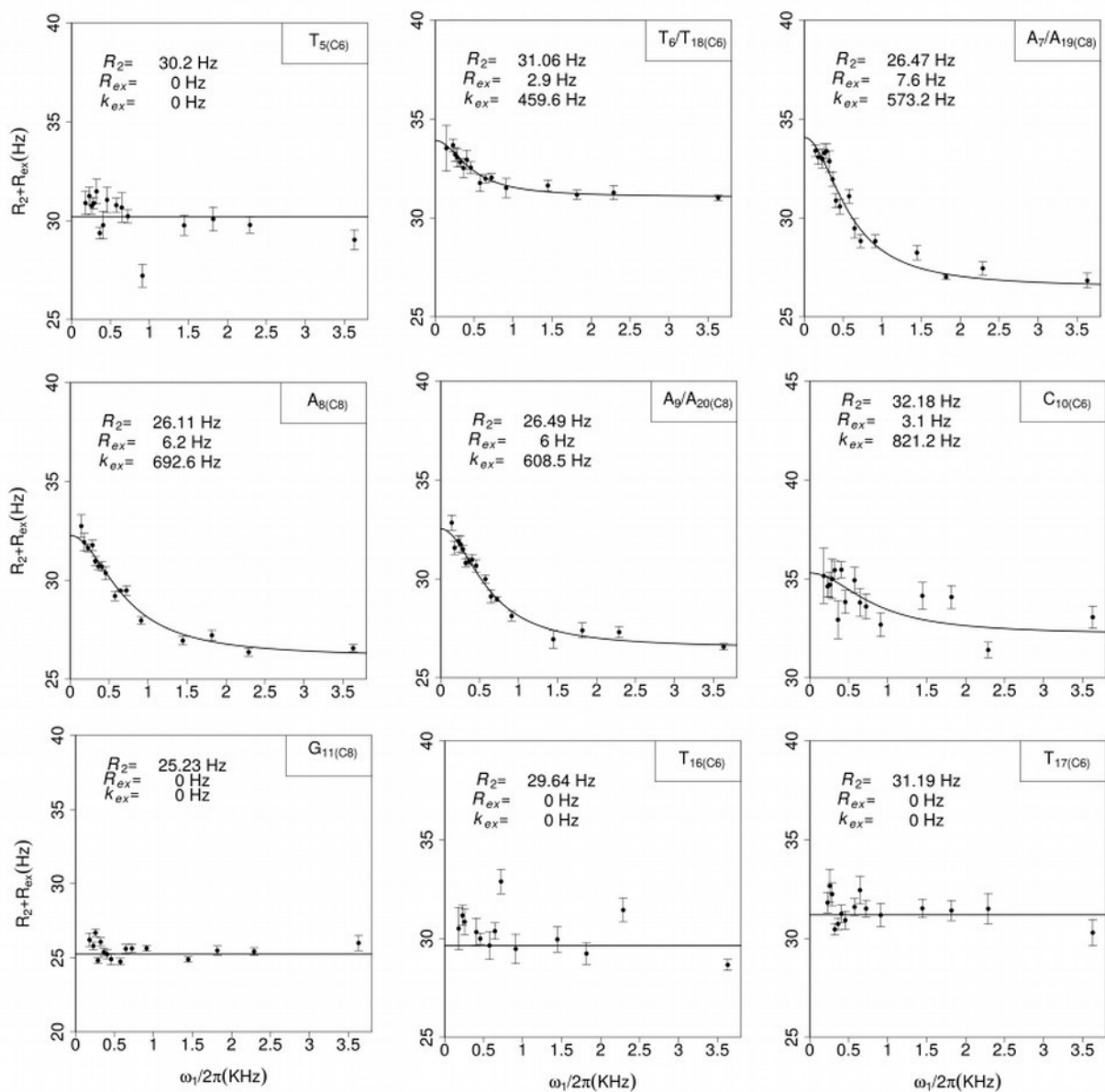


Figure 40: Vitesses de relaxation $R_2 + R_{ex}$ (en Hertz) en fonction de la force de champ ($\omega_1/2\pi$ en Kilo-Hertz) appliquées sur les différents C6/8 des bases aromatiques (légende en haut à gauche). La fréquence de relaxation transversale ($R_2 = 1/T_2$), la constante d'échange (k_{ex}) et la contribution de l'échange à la relaxation (R_{ex}) pour chaque noyau sont mentionnées dans chaque graphique. Les points représentent les mesures expérimentales avec les barres d'erreur correspondantes, les courbes sont données par l'équation 42.

Dans notre étude, par manque de temps, les carbones C1' des sucres n'ont pas été étudiés, mais les données sont disponibles pour les C8, C6 et C2 de plusieurs bases. Les signaux (notamment dans la région C8-H8/C6-H6) des résidus A20 et A9, T6 et T18, et enfin A7 et A19 sont superposés. Il faut toutefois noter que les contextes de séquence sont presque identiques, pour A7/A19 d'une part et

T6/18 d'autre part; dans ce cas il est donc peu probable qu'il soit possible d'observer des différences de déplacements chimiques au niveau des C8/C6 (moins sensibles que les déplacements chimiques des C2 à des variations de séquence très subtiles). Les mouvements se produisent donc vraisemblablement aux deux positions presque symétriques.

Rappelons brièvement ce qui a été publié dans la littérature au sujet des mouvements lents dans les ADN double brin mesurés par dispersion relaxation du ^{13}C . La première étude des mouvements lents sur un ADN double brin a été réalisée en 2001 dans l'équipe de James (Kojima et al. 2001). Ce travail a montré l'existence de mouvements lents sur le carbone C8 ainsi que le proton H8 avec une constante de temps ($1/k_{\text{ex}}$) d'échange de l'ordre de 20 à 300 millisecondes sur la première adénine de la séquence 5'-CAA-3' contenue dans un décimère double-brin.

Beaucoup plus récemment les travaux de l'équipe de Al-Hashimi ont confirmé l'existence de mouvements lents sur les carbones C8 des adénines et des C6 des cytosines des tracts CA_n ($n=2,4,6$ adénines) (Nikolova et al. 2011) (Alvey et al. 2014). De façon plus générale, ces travaux ont montré que dans le cas des transitions Watson-Crick(WC)-Hoogsteen(HG) pour des paires de bases A-T, les carbones C2 des adénines et les C6 des thymine complémentaires ne présentaient pas d'échange chimique dans les gammes perceptibles par les expériences R1 ρ (< 5 millisecondes) alors que des phénomènes d'échange chimique sont observés pour les carbones C8 et C1' des adénines. Concernant les paires de bases G-C, des transitions WC-HG ont également été identifiées; dans ce cas ce sont les C8 et C1' des guanines ainsi que les C6 des cytosines appariées qui présentent des phénomènes d'échanges. Des calculs théoriques de déplacements chimiques du carbone dans les deux conformations WC et HG ont été effectués avec les méthodes de DFT (*Density Functional Theory*) et apparaissent en bon accord avec les données expérimentales de dispersion de relaxation (Nikolova et al., 2011). Il a notamment été bien mis en évidence que la transition WG vers HG dans les paires A-T n'entraîne pas de variation de déplacement chimique pour les carbones C2 des adénines et les C6 des thymine complémentaires alors que les carbones C8 des adénines sont affectés. Le très grand nombre de séquences étudiées dans ces travaux a permis d'étudier les transitions WC-HG dans les paires AT dans de très nombreux contextes de séquence: les variations de déplacements chimiques associés aux transitions (variation des C8 et C1' des adénines ainsi que des C1' des thymine; pas de variation pour les C6 des thymine et les C2 des adénines) apparaissent toujours identiques et semblent donc constituer une signature de la présence de ce type de transition.

4 Conclusion

Dans le Tableau 9, nous avons comparé les résultats que nous avons obtenus à ceux qui sont interprétés comme des transitions WC-HG, issues de l'étude de plusieurs séquences (Nikolova et al. 2011) (Alvey et al. 2014)

Brin 1				Brin2			
séquence	carbone	nos travaux	WC-HG	séquence	carbone	nos travaux	WC-HG
T5	C6	-	-	A20	C8	+	+
					C2	-	-
T6	C6	+	-	A19	C8	+	+
					C2	+	-
A7	C8	+	+	T18	C6	+	-
	C2	+	-				
A8	C8	+	+	T17	C6	-	-
	C2	-	-				
A9	C8	+	+	T16	C6	-	-
	C2	-	-				
C10	C6	+	+	G15	C8	N	+
G11	C8	-	-	C14	C6	N	-

Tableau 9 : Tableau récapitulatif des comparaisons entre les noyaux en mouvement lent issus de nos travaux et leurs équivalent dans les transitions WC-HG des travaux du groupe d'Al-Hashimi (Nikolova et al. 2011) (Alvey et al. 2014). (+) et (-) représentent respectivement l'existence ou non d'un mouvement lent perceptible. Les divergences entre les deux travaux sont indiqué en caractère gras en bleu. La lettre (N) est attribuée pour les noyaux dont la mesure R1 ρ n'est pas encore effectué dans nos travaux.

Les transitions WC vers HG dans les paires AT s'accompagnent de variations de déplacement chimiques des C8 et C1' des adénines et des C1' des sucres des thymines complémentaires. Par contre aucune variation n'est observée sur les C2 des adénines et les C6 des thymines complémentaires.

Dans notre cas, les données collectées sur les trois paires A8-T17 A9-T16 et A20-T5 sont compatibles avec une population excitée dans l'appariement HG. Par contre, de façon tout à fait frappante, les données pour les paires T6-A19 et T18-A7 ne peuvent pas être interprétées comme des transitions WC-HG puisque des mouvements lents sont détectés non seulement sur les carbones C8 mais aussi sur les carbones C2 et C6 des thymines. Ces conclusions sont résumées dans le schéma suivant, avec en rouge les paires de bases fortement soupçonnées d'être en équilibre entre des états WC et HG, et en vert le bloc central TA.TA qui est aussi l'objet d'une transitions entre deux états se déroulant toujours dans le domaine de la milliseconde mais d'un type nouveau, différent d'un équilibre conformationnel WC ↔ HG.

5'-CCGCTTAACGC-3'

3'-GGCGAATTTGCG-5'

Il est clair que d'autres informations, notamment celles que l'on peut extraire des expériences de dispersion de relaxation "*Off-resonance*" seront nécessaires pour déterminer le type de mouvement impliqué.

CONCLUSION ET PERSPECTIVES

Dans le cadre de cette thèse nous avons étudié la dynamique de l'ADN-B à deux échelles de temps différentes. Concernant les mouvement rapides (picoseconde-nanoseconde), nous nous sommes concentrés sur l'équilibre BI \leftrightarrow BII des groupements phosphates, dans la continuité d'une thématique développée au laboratoire. Dans cette partie, nous avons utilisé des expériences de RMN de routine et des simulations de dynamique moléculaire. Ces approches ont été réalisées sur quatre dodécamères dont les séquences sont reliées à un ADN nommé « séquence 601 » qui est connu pour favoriser la formation du nucléosome.

Nous nous sommes ensuite tournés vers des échanges conformationnels plus lents (microseconde-milliseconde), encore très peu abordés sur des ADN en double hélice. Dans la première étape de cette thèse, nous avons décelé des signes de mouvements lents sur l'un des dodécamères. Nous avons approfondi nos investigations en marquant cet oligomère et en l'étudiant par des expériences de RMN de dispersion de relaxation que nous avons introduites et mises au point au laboratoire.

1 Mouvements rapides

Grâce aux données RMN collectées sur les quatre dodécamères, nous avons d'abord validé l'effet de séquence sur l'équilibre BI \leftrightarrow BII et définitivement confirmé les prédictions de l'échelle TRX, publiée en 2010 (Heddi et al. 2010), qui associe une population de BII spécifique à chaque dinucléotidique constituant l'ADN-B en solution. L'appellation TRX a également été pleinement justifiée. En effet le terme « TRX » fait référence aux corrélations entre les conformations BI et BII des groupements phosphate et les paramètres hélicoïdaux de *Twist*, *Roll* et *X-disp* (déplacement des paires de bases), initialement observées à partir de l'analyse de structures cristallographiques. En montrant qu'il existe des corrélations entre δP , distances internucléotides et ΔRDC , et en interprétant ces données expérimentales en termes structuraux, nous avons pu mettre en évidence pour la première fois en solution la mécanique intrinsèque de l'ADN : les *Twist*, *Roll* et *Slide* sont sensibles à l'équilibre BI \leftrightarrow BII, ainsi que la taille du petit sillon, reliée au *X-disp*.

Enfin, nous avons testé les performances de Parmbsc0_{OL1} et CHARMM36, deux très récents champs de force mis au point pour améliorer la représentation des mouvements des groupements

phosphate. En comparant des dynamiques aux données RMN, il s'est avéré que CHARMM36 en particulier génère des structures beaucoup plus satisfaisantes et réalistes que les champs de force « parents » ou autres. Cette amélioration nous a permis de d'extraire des informations inédites sur les équilibres BI ↔ BII dans des pas dinucléotidiques complémentaires.

Ces travaux ouvrent sur au moins deux perspectives, dont l'une est en cours de réalisation. Ces ouvertures sont en fait basées sur des simulations avec CHARMM36, qui nous apparaît comme le plus réaliste des champs de force actuels.

La première chose est d'obtenir des dynamiques sur les quatre dodécamères, mais cette fois en utilisant des contraintes RMN, distances interatomiques et Δ RDC, qui pourraient bénéficier des corrélations que nous avons mises en évidence entre ces observables et les δ P. En effet, tous les δ P ont été mesurés le long des quatre dodécamères, ce qui n'est pas le cas pour les distances interatomiques et les RDC qui sont donc inégalement répartis le long des séquences. Dans des dynamiques moléculaires nous pouvons imaginer restituer les contraintes manquantes en les extrapolant à partir des équations de corrélation qui donnent les distances interatomiques ou les Δ RDC en fonction des δ P mesurés. Ce type de simulations nous permettrait de mesurer à quel point l'ajout de données « expérimentales » (directement ou indirectement obtenues) modifie les propriétés structurales des oligomères modélisés sans contraintes. Cette question n'est pas triviale dans la mesure où CHARMM36 rend d'emblée assez bien compte des équilibres BI ↔ BII. Mais quelque soit la réponse, nous disposerions pour les quatre dodécamères de structures solidement étayées par la RMN, que nous pourrions ensuite comparer à leurs équivalents dans le nucléosome.

Toujours en relation avec le nucléosome, nous avons vu que l'échelle TRX a été utilisée pour caractériser les séquences les plus affines du cœur d'histone, soit dans ce travail (chapitre I), soit dans un travail précédent (Heddi et al. 2010). En bref, les séquences les plus aptes à former des nucléosomes semblent être organisées de telle manière qu'elles alternent des régions à faible et forte propension BII. Mais qu'en est-il dans le nucléosome ? Pour répondre à cette question, Ahmad Elbahnsi réalise actuellement avec CHARMM36 des dynamiques moléculaires de nucléosomes contenant différentes séquences dont la séquence 601. Les résultats préliminaires sont très encourageants puisqu'ils montrent que dans le nucléosome, il apparaît une périodicité caractérisée par des régions stabilisées soit en BI soit en BII et ce, quelque soit la séquence. Cette étude prouverait donc que le cœur d'histone imposerait cette périodicité et que les séquences possédant les propriétés intrinsèques qui vont dans ce sens seraient les plus aptes à former les nucléosomes

puisqu'elles minimiserait le coût de déformation.

2 Mouvements lents

L'observation des largeurs de raies puis les expériences de dispersion de relaxation ont montré l'existence d'échanges conformationnels à l'échelle de la milliseconde sur le bloc TTAAA.TTTAA, qui appartient à l'Oligo 3 dont la séquence complète est CCGCTTAAACGC.GCGTTTAAGCGG. Il reste maintenant à caractériser ces échanges plus en détail. De fait, il s'agit de collecter le plus d'informations possibles sur l'état conformationnel minoritaire, puisque nous savons que l'état majoritaire est la double hélice B. Les données actuelles pourraient ainsi être complétées par des études à d'autres températures pour tenter de séparer les résonances des C8 de deux adénines, actuellement difficilement exploitables parce que superposées. L'étude par dispersion de relaxation des carbones des sucres (notamment des C1') et des azotes des bases serait également très importante pour mieux connaître l'identité de l'ensemble des noyaux impliqués dans des échanges conformationnels. Des expériences R1ρ *Off-resonance* seront ensuite essentielles pour pouvoir déterminer les déplacements chimiques des états minoritaires. De plus, ces expériences permettront d'obtenir les pourcentages de l'état minoritaire de l'échange et de pouvoir ainsi calculer l'énergie libre de l'échange. Enfin, une fois que les déplacements chimiques des états minoritaires auront été déterminés, il faudra tenter d'identifier les structures compatibles avec ces données soit par des recherches dans les bases de données RMN, soit par des simulations type modes normaux, soit encore par des calculs de chimie quantique sur des modèles théoriques.

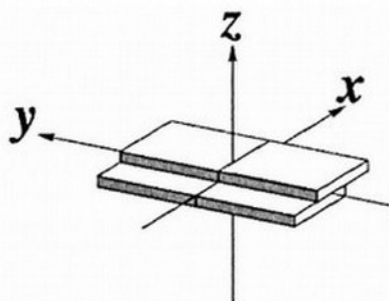
Mais déjà tels qu'ils sont, les résultats que nous avons obtenus sont intéressants. L'originalité n'est pas dans l'existence de mouvements lents, qui ont été détectés pour la première fois en observant les élargissements des raies des H2 des adénines dans des dinucléotides TpA (J.-F. Lefevre, Lane, et Jardetzky 1985) (J. F. Lefevre, Lane, et Jardetzky 1988) (Kennedy et al. 1993) (Lingbeck et al. 1996) Il s'est ensuite trouvé que ces élargissements, bien que se produisant quasi-systématiquement sur les pas TpA quelque soit leur contexte tétramérique, étaient les plus forts dans des fragments YTAA (McAteer et Kennedy 2000) . Les élargissements que nous avons observés sur les pas TpA complémentaires appartenant à un contexte TTAA sont donc conformes à ces observations.

Nos expériences de dispersion de relaxation ont confirmé l'existence de mouvements lents sur les TpA, mais ont aussi révélé que l'ensemble du bloc TTAAA.TTTAA était concerné. De telles

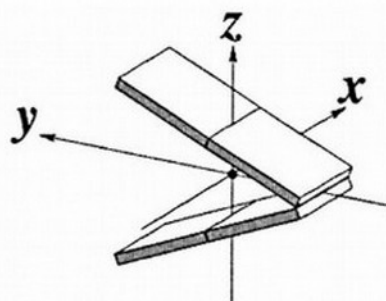
régions couvrant plusieurs paires de base et soumises à des transitions identifiées comme des équilibres entre appariements Watson-Crick et appariements Hoogsteen ont récemment été mises en évidence (Alvey et al. 2014) . Mais là où nos données sont vraiment originales, c'est qu'elles suggèrent fortement qu'il existe deux types d'équilibre dans le bloc TTAAA.TTTAA. Le premier équilibre concerne les paires de bases voisines des TpA (A:T soulignées) et ressemble à celui déjà identifié (Watson-Crick ↔ Hoogsteen) ; le deuxième équilibre se produit sur les deux pas TpA (en rouge) qui se font face et n'a pas les mêmes caractéristiques que le premier. Il correspond donc à une transition de la double hélice vers une forme minoritaire encore inconnue. A notre connaissance, c'est la première fois que deux mouvements différents sont identifiés au sein d'un même fragment.

D'un point de vue plus biologique, il est aussi très important de mieux identifier ces équilibres. L'élément TTAAA.TTTAA est en effet connu pour être un facteur clé dans l'ancrage du cœur d'histone sur l'ADN, plus précisément dans la fixation des histones H3 et H4 (Chua et al. 2012). Quand il est muté, l'affinité de la séquence 601 décroît. A l'inverse, quand il est inséré dans une séquence d'affinité modeste, celle-ci augmente. Dans le nucléosome, une arginine s'insère dans le petit sillon de TTAAA.TTTAA, cette interaction étant verrouillée par une leucine et une proline. Nous avons vu dans le chapitre I qu'à l'état libre, le petit sillon de cet élément d'ADN était particulièrement étroit, une caractéristique qui, quand elle est associée à des plateaux de bases A:T, est connue pour générer un fort potentiel électronégatif très attractif pour l'arginine. Mais il est aussi possible que les équilibres conformationnels lents que nous avons mis en évidence puissent intervenir dans la formation de ce point d'ancrage. Cette hypothèse demande bien sûr à être prouvée, mais elle ouvre des perspectives tout à fait nouvelles dans le domaine des complexes ADN-protéines et des mécanismes qui sous-tendent leur formation.

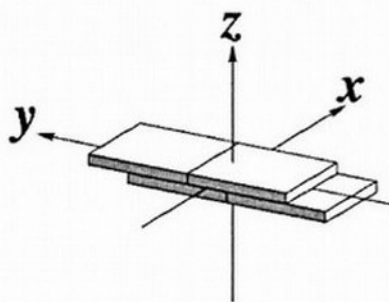
ANNEXE A



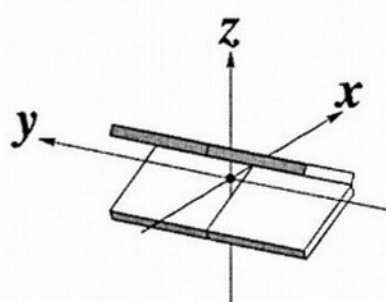
Shift



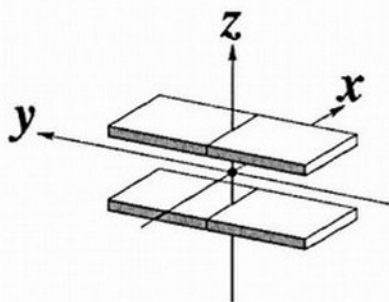
Tilt



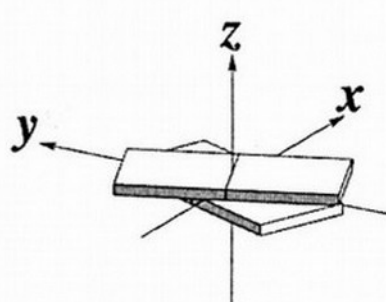
Slide



Roll



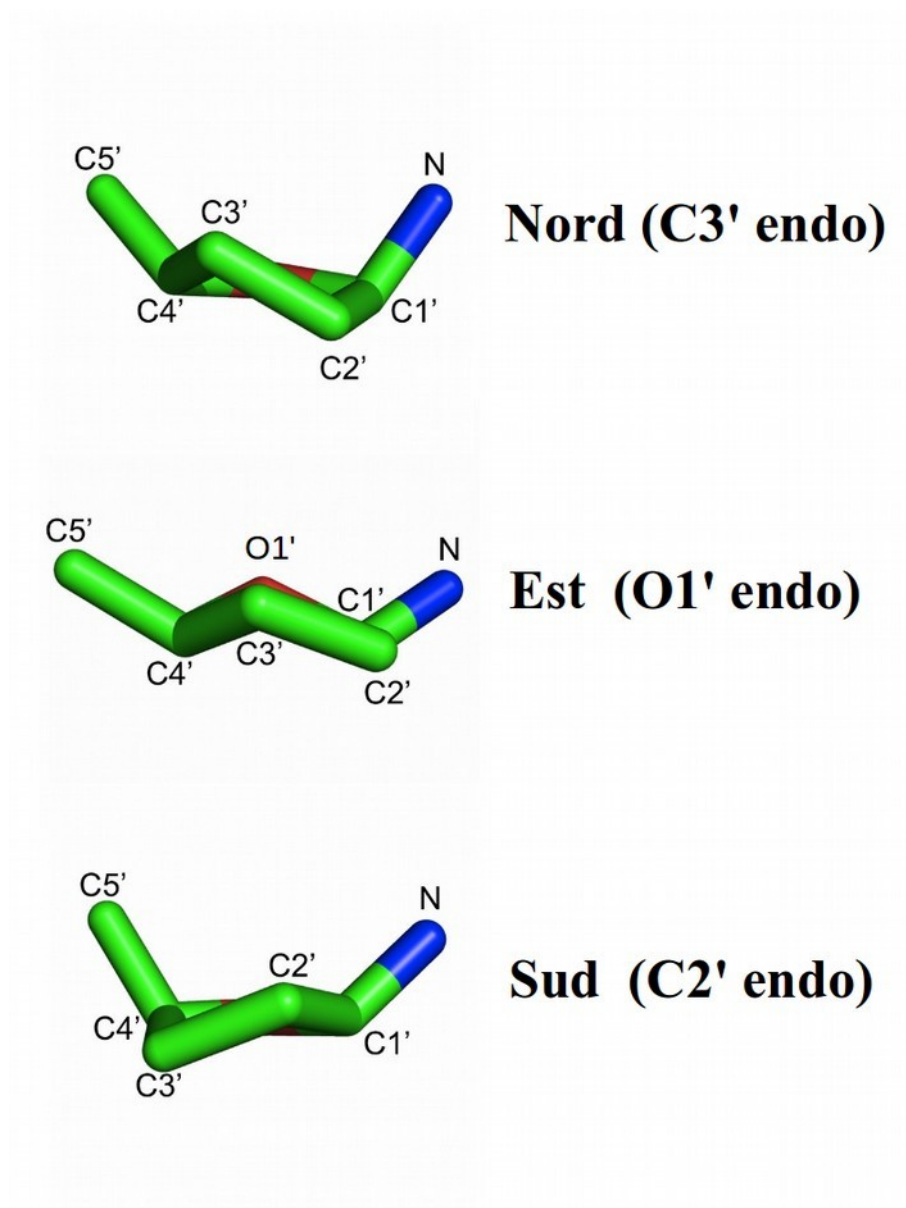
Rise



Twist

Annexe A : Représentation schématique des paramètres hélicoïdaux relatifs au paires de bases consécutives

ANNEXE B



Annexe B : Représentation des trois conformations principales des désoxyriboses : Nord, Est et Sud.

BIBLIOGRAPHIE

- Abi-Ghanem, Joséphine, Brahim Heddi, Nicolas Foloppe, et Brigitte Hartmann. 2009. « DNA Structures from Phosphate Chemical Shifts ». *Nucleic Acids Research* 38 (3): e18-e18. doi:10.1093/nar/gkp1061.
- Alvey, Heidi S., Federico L. Gottardo, Evgenia N. Nikolova, et Hashim M. Al-Hashimi. 2014. « Widespread Transient Hoogsteen Base Pairs in Canonical Duplex DNA with Variable Energetics ». *Nature Communications* 5 (septembre). doi:10.1038/ncomms5786.
- Baldwin, Andrew J., et Lewis E. Kay. 2009. « NMR Spectroscopy Brings Invisible Protein States into Focus ». *Nature Chemical Biology* 5 (11): 808-14. doi:10.1038/nchembio.238.
- Bax, Ad, et Donald G Davis. 1985. « Practical aspects of two-dimensional transverse NOE spectroscopy ». *Journal of Magnetic Resonance* (1969) 63 (1): 207-13. doi:10.1016/0022-2364(85)90171-4.
- Bax, Ad, Georg Kontaxis, et Nico Tjandra. 2001. « [8] - Dipolar Couplings in Macromolecular Structure Determination ». In *Methods in Enzymology*, édité par Volker Dötsch and Uli Schmitz Thomas L. James, 339:127-74. Nuclear Magnetic Resonance of Biological Macromolecules - Part B. Academic Press. <http://www.sciencedirect.com/science/article/pii/S0076687901393138>.
- Bendel, P., et T. L. James. 1983. « Structural and Dynamic Differences between Supercoiled and Linear DNA from Proton NMR ». *Proceedings of the National Academy of Sciences* 80 (11): 3284-86.
- Berman, Helen M. 1997. « Crystal Studies of B-DNA: The Answers and the Questions ». *Biopolymers* 44 (1): 23-44. doi:10.1002/(SICI)1097-0282(1997)44:1<23::AID-BIP3>3.0.CO;2-1.
- Bertrand, H.-O., S. Fermandjian, T. Ha-Duong, et B. Hartmann. 1998. « Flexibility of the B-DNA Backbone: Effects of Local and Neighbouring Sequences on Pyrimidine-Purine Steps ». *Nucleic Acids Research* 26 (5): 1261-67. doi:10.1093/nar/26.5.1261.
- Beveridge, David L., Gabriela Barreiro, K. Suzie Byun, David A. Case, Thomas E. Cheatham III, Surjit B. Dixit, Emmanuel Giudice, et al. 2004. « Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. I. Research Design and Results on d(CpG) Steps ». *Biophysical Journal* 87 (6): 3799-3813. doi:10.1529/biophysj.104.045252.
- Boisbouvier, Jerome, Zhengrong Wu, Akira Ono, Masatsune Kainosho, et Ad Bax. 2003. « Rotational Diffusion Tensor of Nucleic Acids from ¹³C NMR Relaxation ». *Journal of Biomolecular NMR* 27 (2): 133-42. doi:10.1023/A:1024931619957.
- Borer, Philip N., Steven R. LaPlante, Anil Kumar, Nilo Zanatta, Amy Martin, Anna Hakkinen, et George C. Levy. 1994. « ¹³C-NMR Relaxation in Three DNA Oligonucleotide Duplexes: Model-Free Analysis of Internal and Overall Motion ». *Biochemistry* 33 (9): 2441-50. doi:10.1021/bi00175a012.
- Bothe, Jameson R., Ky Lowenhaupt, et Hashim M. Al-Hashimi. 2011. « Sequence-Specific B-DNA Flexibility Modulates Z-DNA Formation ». *Journal of the American Chemical Society* 133 (7): 2016-18. doi:10.1021/ja1073068.
- Cheatham, Thomas E., Piotr Cieplak, et Peter A. Kollman. 1999. « A Modified Version of the Cornell *et Al.* Force Field with Improved Sugar Pucker Phases and Helical Repeat ». *Journal of Biomolecular Structure and Dynamics* 16 (4): 845-62. doi:10.1080/07391102.1999.10508297.
- Chiarparin, Elisabetta, Philippe Pelupessy, et Geoffray Bodenhausen. 1998. « Selective cross-polarization in solution state NMR ». *Molecular Physics* 95 (5): 759-67. doi:10.1080/002689798166396.
- Chou, Shan-Ho, Jya-Wei Cheng, et Brian R. Reid. 1992. « Solution Structure of [d(ATGAGCGAATA)]₂ ». *Journal of Molecular Biology* 228 (1): 138-55. doi:10.1016/0022-2836(92)90497-8.
- Chua, Eugene Y. D., Dileep Vasudevan, Gabriela E. Davey, Bin Wu, et Curt A. Davey. 2012. « The Mechanics behind DNA Sequence-Dependent Properties of the Nucleosome ». *Nucleic Acids Research* 40 (13): 6338-52. doi:10.1093/nar/gks261.
- Clore, G. Marius, Attila Szabo, Ad Bax, Lewis E. Kay, Paul C. Driscoll, et Angela M. Gronenborn. 1990.

- « Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins ». *Journal of the American Chemical Society* 112 (12): 4989-91. doi:10.1021/ja00168a070.
- Cornell, Wendy D., Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, et Peter A. Kollman. 1996. « A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules J. Am. Chem. Soc. 1995, 117, 5179–5197 ». *Journal of the American Chemical Society* 118 (9): 2309-2309. doi:10.1021/ja955032e.
- Darden, Tom, Darrin York, et Lee Pedersen. 1993. « Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems ». *The Journal of Chemical Physics* 98 (12): 10089-92. doi:10.1063/1.464397.
- Davis, D. G., M. E. Perlman, et R. E. London. 1994. « Direct Measurements of the Dissociation-Rate Constant for Inhibitor-Enzyme Complexes via the T1 ρ and T2 (CPMG) Methods ». *Journal of Magnetic Resonance, Series B* 104 (3): 266-75. doi:10.1006/jmrb.1994.1084.
- Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, et A. Bax. 1995. « NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes ». *Journal of Biomolecular NMR* 6 (3): 277-93.
- Dixit, Surjit B., David L. Beveridge, David A. Case, Thomas E. Cheatham 3rd, Emmanuel Giudice, Filip Lankas, Richard Lavery, et al. 2005. « Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. II: Sequence Context Effects on the Dynamical Structures of the 10 Unique Dinucleotide Steps ». *Biophysical Journal* 89 (6): 3721-40. doi:10.1529/biophysj.105.067397.
- Djuranovic, D., et B. Hartmann. 2003. « Conformational Characteristics and Correlations in Crystal Structures of Nucleic Acid Oligonucleotides: Evidence for Sub-states ». *Journal of Biomolecular Structure and Dynamics* 20 (6): 771-88. doi:10.1080/07391102.2003.10506894.
- Djuranovic, D., et B. Hartmann. 2004. « DNA Fine Structure and Dynamics in Crystals and in Solution: The Impact of BI/BII Backbone Conformations ». *Biopolymers* 73 (3): 356-68. doi:10.1002/bip.10528.
- Drew, H. R., R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, et R. E. Dickerson. 1981. « Structure of a B-DNA Dodecamer: Conformation and Dynamics ». *Proceedings of the National Academy of Sciences* 78 (4): 2179-83.
- Duchardt, Elke, Lennart Nilsson, et Jürgen Schleucher. 2008. « Cytosine ribose flexibility in DNA: a combined NMR ^{13}C spin relaxation and molecular dynamics simulation study ». *Nucleic Acids Research* 36 (12): 4211-19. doi:10.1093/nar/gkn375.
- Early, Thomas A., David R. Kearns, Wolfgang Hillen, et Robert D. Wells. 1981a. « A 300- and 600-MHz proton nuclear magnetic resonance investigation of a 12 base pair deoxyribonucleic acid restriction fragment: relaxation behavior of the low-field resonances in water ». *Biochemistry* 20 (13): 3756-64. doi:10.1021/bi00516a014.
- Early, Thomas A., David R. Kearns, Wolfgang Hillen, et Robert D. Wells. 1981b. « A 300 MHz proton nuclear magnetic resonance investigation of deoxyribonucleic acid restriction fragments: dynamic properties ». *Biochemistry* 20 (13): 3764-69. doi:10.1021/bi00516a015.
- El antri, Said, Patrick Bittoun, Olivier Mauffret, Monique Monnot, Odile Convert, Elie Lescot, et Serge Fermandjian. 1993. « Effect of distortions in the phosphate backbone conformation of six related octanucleotide duplexes on CD and phosphorus-31 NMR spectra ». *Biochemistry* 32 (28): 7079-88. doi:10.1021/bi00079a003.
- Feigon, Juli, Werner Leupin, William A. Denny, et David R. Kearns. 1983. « Two-dimensional proton nuclear magnetic resonance investigation of the synthetic deoxyribonucleic acid decamer d(ATATCGATAT) ». *Biochemistry* 22 (25): 5943-51. doi:10.1021/bi00294a038.
- Foloppe, Nicolas, et Jr. MacKerell Alexander D. 2000. « All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data ». *Journal of Computational Chemistry* 21 (2): 86-104. doi:10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G.
- Fratini, A. V., M. L. Kopka, H. R. Drew, et R. E. Dickerson. 1982. « Reversible Bending and Helix Geometry in a B-DNA Dodecamer: CGCGAATTBrCGCG. » *Journal of Biological Chemistry* 257 (24): 14686-707.
- Galindo-Murillo, Rodrigo, Daniel R. Roe, et Thomas E. Cheatham III. 2015. « Convergence and

- reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC) ». *Biochimica et Biophysica Acta (BBA) - General Subjects*, Recent developments of molecular dynamics, 1850 (5): 1041-58. doi:10.1016/j.bbagen.2014.09.007.
- Geen, Helen, et Ray Freeman. 1991. « Band-selective radiofrequency pulses ». *Journal of Magnetic Resonance* (1969) 93 (1): 93-141. doi:10.1016/0022-2364(91)90034-Q.
- Giessner-Prettre, Claude, et Bernard Pullman. 1970. « Intermolecular nuclear shielding values for protons of purines and flavins ». *Journal of Theoretical Biology* 27 (1): 87-95. doi:10.1016/0022-5193(70)90130-X.
- Gorenstein, David G. 1984. *Phosphorous-31 NMR: Principles and Applications*. Academic Press.
- Gorenstein, David G. 1987. « Stereoelectronic effects in biomolecules ». *Chemical Reviews* 87 (5): 1047-77. doi:10.1021/cr00081a009.
- Gorenstein, D. G. 1992. « ³¹P NMR of DNA ». *Methods in enzymology* 211: 254-86.
- Gorin, Andrey A., Victor B. Zhurkin, et Wilma K. 1995. « B-DNA Twisting Correlates with Base-pair Morphology ». *Journal of Molecular Biology* 247 (1): 34-48. doi:10.1006/jmbi.1994.0120.
- Hansen, Alexandar L., Evgenia N. Nikolova, Anette Casiano-Negroni, et Hashim M. Al-Hashimi. 2009. « Extending the Range of Microsecond-to-Millisecond Chemical Exchange Detected in Labeled and Unlabeled Nucleic Acids by Selective Carbon R1p NMR Spectroscopy ». *Journal of the American Chemical Society* 131 (11): 3818-19. doi:10.1021/ja8091399.
- Hare, Dennis R., et Brian R. Reid. 1986. « Three-dimensional structure of a DNA hairpin in solution: two-dimensional NMR studies and distance geometry calculations on d(CGCGTTTTCGCG) ». *Biochemistry* 25 (18): 5341-50. doi:10.1021/bi00366a053.
- Hart, Katarina, Nicolas Foloppe, Christopher M. Baker, Elizabeth J. Denning, Lennart Nilsson, et Alexander D. MacKerell. 2012. « Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium ». *Journal of Chemical Theory and Computation* 8 (1): 348-62. doi:10.1021/ct200723y.
- Havel, Timothy, et Kurt Wüthrich. 1984. « A Distance Geometry Program for Determining the Structures of Small Proteins and Other Macromolecules from Nuclear Magnetic Resonance Measurements of intramolecular ¹H-¹H Proximities in Solution ». *Bulletin of Mathematical Biology* 46 (4): 673-98. doi:10.1007/BF02459510.
- Heddi, Brahim, Nicolas Foloppe, Nadia Bouchemal, Edith Hantz, et Brigitte Hartmann. 2006. « Quantification of DNA BI/BII Backbone States in Solution. Implications for DNA Overall Structure and Recognition ». *Journal of the American Chemical Society* 128 (28): 9170-77. doi:10.1021/ja061686j.
- Heddi, Brahim, Nicolas Foloppe, Christophe Oguey, et Brigitte Hartmann. 2008. « Importance of Accurate DNA Structures in Solution: The Jun-Fos Model ». *Journal of Molecular Biology* 382 (4): 956-70. doi:10.1016/j.jmb.2008.07.047.
- Heddi, Brahim, Christophe Oguey, Christophe Lavelle, Nicolas Foloppe, et Brigitte Hartmann. 2010. « Intrinsic Flexibility of B-DNA: The Experimental TRX Scale ». *Nucleic Acids Research*, novembre, gkp962. doi:10.1093/nar/gkp962.
- Heinemann, Udo, Claudia Alings, et Michael Hahn. 1994. « Crystallographic studies of DNA helix structure ». *Biophysical Chemistry*, 11th International Biophysics Congress, 50 (1-2): 157-67. doi:10.1016/0301-4622(94)85028-3.
- Hogan, M. E., et O. Jardetzky. 1980. « Internal Motions in Deoxyribonucleic Acid II ». *Biochemistry* 19 (15): 3460-68.
- Holbrook, S. R., R. E. Dickerson, et S.-H. Kim. 1985. « Anisotropic thermal-parameter refinement of the DNA dodecamer CGCGAATTCGCG by the segmented rigid-body method ». *Acta Crystallographica Section B Structural Science* 41 (4): 255-62. doi:10.1107/S0108768185002087.
- Irvine, Doug, Craig Tuerk, et Larry Gold. 1991. « Selexion: Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis ». *Journal of Molecular Biology* 222 (3): 739-61. doi:10.1016/0022-2836(91)90509-5.
- Isaacs, Richard J, et H. Peter Spielmann. 2001. « NMR evidence for mechanical coupling of phosphate BI-BII transitions with deoxyribose conformational exchange in DNA1 ». *Journal of Molecular Biology* 311 (1): 149-60. doi:10.1006/jmbi.2001.4855.

- James G Kempf, Ju-yeon Jung. 2003. « Off-Resonance TROSY ($R_{1\rho} - R_1$) for Quantitation of Fast Exchange Processes in Large Proteins ». *Journal of the American Chemical Society* 125 (40): 12064-65. doi:10.1021/ja037101s.
- Karplus, Martin. 1959. « Contact Electron-Spin Coupling of Nuclear Magnetic Moments ». *The Journal of Chemical Physics* 30 (1): 11-15. doi:10.1063/1.1729860.
- Kennedy, Michael A., Sirkku T. Nuutero, Jeffery T. Davis, Gary P. Drobny, et Brian R. Reid. 1993. « Mobility at the TpA cleavage site in the T3A3-containing AhaIII and PmeI restriction sequences ». *Biochemistry* 32 (31): 8022-35. doi:10.1021/bi00082a025.
- Klevan, Leonard, Ian M. Armitage, et Donald M. Crothers. 1979. « ^{31}P NMR Studies of the Solution Structure and Dynamics of Nucleosomes and DNA ». *Nucleic Acids Research* 6 (4): 1607-16. doi:10.1093/nar/6.4.1607.
- Kojima, Chojiro, Nikolai B. Ulyanov, Masatsune Kainosho, et Thomas L. James. 2001. « Slow Motion in the CAA·TTG Sequence of a DNA Decamer Duplex Studied by NMR† ». *Biochemistry* 40 (24): 7239-46. doi:10.1021/bi0103260.
- Korzhnev, Dmitry M., Vladislav Yu. Orekhov, et Lewis E. Kay. 2005. « Off-Resonance $R_{1\rho}$ NMR Studies of Exchange Dynamics in Proteins with Low Spin-Lock Fields: An Application to a Fyn SH3 Domain ». *Journal of the American Chemical Society* 127 (2): 713-21. doi:10.1021/ja0446855.
- Kuntz, Irwin D., Jeffrey M. Blaney, Stuart J. Oatley, Robert Langridge, et Thomas E. Ferrin. 1982. « A geometric approach to macromolecule-ligand interactions ». *Journal of Molecular Biology* 161 (2): 269-88. doi:10.1016/0022-2836(82)90153-X.
- Lankhorst, Peter P., Cornells A. G. Haasnoot, Cornells Erkelens, et Cornells Altona. 1984. « Carbon-13 NMR in Conformational Analysis of Nucleic Acid Fragments 2. A Reparametrization of the Karplus Equation for Vicinal NMR Coupling Constants in CCOP and HCOP Fragments ». *Journal of Biomolecular Structure and Dynamics* 1 (6): 1387-1405. doi:10.1080/07391102.1984.10507527.
- Lavery, Richard. 1988. « DNA Flexibility Under Control: The Jumna Algorithm and Its Application to BZ Junctions ». In *Unusual DNA Structures*, édité par Robert D. Wells et Stephen C. Harvey, 189-206. Springer New York. http://link.springer.com/chapter/10.1007/978-1-4612-3800-3_11.
- Lavery, R., K. Zakrzewska, et H. Sklenar. 1995. « JUMNA (junction minimisation of nucleic acids) ». *Computer Physics Communications* 91 (1-3): 135-58. doi:10.1016/0010-4655(95)00046-I.
- Lefebvre, A., O. Mauffret, B. Hartmann, E. Lescot, et S. Femandjian. 1995. « Structural Behavior of the CpG Step in Two Related Oligonucleotides Reflects Its Malleability in Solution ». *Biochemistry* 34 (37): 12019-28. doi:10.1021/bi00037a045.
- Lefebvre, A., O. Mauffret, E. Lescot, B. Hartmann, et S. Femandjian. 1996. « Solution Structure of the CpG Containing d(CTTCGAAG)₂ Oligonucleotide: NMR Data and Energy Calculations Are Compatible with a BI/BII Equilibrium at CpG ». *Biochemistry* 35 (38): 12560-69. doi:10.1021/bi9606298.
- Lefevre, Jean-François, Andrew N. Lane, et Oleg Jardetzky. 1985. « A Temperature Dependent Transition in the Pribnow Box of the Trp Promoter ». *FEBS Letters* 190 (1): 37-40. doi:10.1016/0014-5793(85)80422-1.
- Lefevre, Jean Francois, Andrew N. Lane, et Oleg Jardetzky. 1988. « A description of conformational transitions in the Pribnow box of the trp promoter of Escherichia coli ». *Biochemistry* 27 (4): 1086-94. doi:10.1021/bi00404a002.
- Levitt, M. 1983. « Computer Simulation of DNA Double-Helix Dynamics ». *Cold Spring Harbor Symposia on Quantitative Biology* 47 (janvier): 251-62. doi:10.1101/SQB.1983.047.01.030.
- Lingbeck, Jody, Mark G. Kubinec, Julie Miller, Brian R. Reid, Gary P. Drobny, et Michael A. Kennedy. 1996. « Effect of Adenine Methylation on the Structure and Dynamics of TpA Steps in DNA: NMR Structure Determination of [d(CGAGGTTTAAACCTCG)]₂ and Its A9-Methylated Derivative at 750 MHz ». *Biochemistry* 35 (3): 719-34. doi:10.1021/bi951364k.
- Lipari, Giovanni, et Attila Szabo. 1981. « Nuclear magnetic resonance relaxation in nucleic acid fragments: models for internal motion ». *Biochemistry* 20 (21): 6250-56. doi:10.1021/bi00524a053.
- Lipari, Giovanni, et Attila Szabo. 1982. « Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity ». *Journal of the American Chemical Society* 104 (17): 4546-59. doi:10.1021/ja00381a009.
- Luginbühl, Peter, et Kurt Wüthrich. 2002. « Semi-classical nuclear spin relaxation theory revisited for use

- with biological macromolecules ». *Progress in Nuclear Magnetic Resonance Spectroscopy* 40 (3): 199-247. doi:10.1016/S0079-6565(01)00043-7.
- Lundström, Patrik, et Mikael Akke. 2005. « Off-Resonance Rotating-Frame Amide Proton Spin Relaxation Experiments Measuring Microsecond Chemical Exchange in Proteins ». *Journal of Biomolecular NMR* 32 (2): 163-73. doi:10.1007/s10858-005-5027-3.
- MacDonald, Douglas, Kristina Herbert, Xiaolin Zhang, Thomas Polgruto, et Ponzy Lu. 2001. « Solution structure of an A-tract DNA bend1 ». *Journal of Molecular Biology* 306 (5): 1081-98. doi:10.1006/jmbi.2001.4447.
- MacKerell, Alexander D., Nilesh Banavali, et Nicolas Foloppe. 2000. « Development and Current Status of the CHARMM Force Field for Nucleic Acids ». *Biopolymers* 56 (4): 257-65. doi:10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W.
- MacKerell, Alexander D., Nilesh Banavali, et Nicolas Foloppe. 2001. « Development and Current Status of the CHARMM Force Field for Nucleic Acids ». *Biopolymers* 56 (4): 257-65. doi:10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W.
- Maehigashi, Tatsuya, Chiaolong Hsiao, Kristen Kruger Woods, Tinoush Moulaei, Nicholas V. Hud, et Loren Dean Williams. 2012. « B-DNA Structure Is Intrinsically Polymorphic: Even at the Level of Base Pair Positions ». *Nucleic Acids Research* 40 (8): 3714-22. doi:10.1093/nar/gkr1168.
- Makde, Ravindra D., Joseph R. England, Hemant P. Yennawar, et Song Tan. 2010. « Structure of RCC1 Chromatin Factor Bound to the Nucleosome Core Particle ». *Nature* 467 (7315): 562-66. doi:10.1038/nature09321.
- Massi, Francesca, Eric Johnson, Chunyu Wang, Mark Rance, et Arthur G. Palmer. 2004. « NMR Rotating-Frame Relaxation with Weak Radio Frequency Fields ». *Journal of the American Chemical Society* 126 (7): 2247-56. doi:10.1021/ja038721w.
- Mauffret, O., B. Hartmann, O. Convert, R. Lavery, et S. Femandjian. 1992. « The fine structure of two DNA dodecamers containing the cAMP responsive element sequence and its inverse: Nuclear magnetic resonance and molecular simulation studies ». *Journal of Molecular Biology* 227 (3): 852-75. doi:10.1016/0022-2836(92)90227-B.
- Mauffret, Olivier, Georges Tevanian, et Serge Femandjian. 2002. « Residual Dipolar Coupling Constants and Structure Determination of Large DNA Duplexes ». *Journal of Biomolecular NMR* 24 (4): 317-28. doi:10.1023/A:1021645131882.
- McAteer, Kathleen, et Michael A. Kennedy. 2000. « NMR Evidence for Base Dynamics at all TpA Steps in DNA ». *Journal of Biomolecular Structure and Dynamics* 17 (6): 1001-9. doi:10.1080/07391102.2000.10506588.
- McCammon, J. Andrew, Bruce R. Gelin, et Martin Karplus. 1977. « Dynamics of folded proteins ». *Nature* 267 (5612): 585-90. doi:10.1038/267585a0.
- McConnell, Harden M. 1958. « Reaction Rates by Nuclear Magnetic Resonance ». *The Journal of Chemical Physics* 28 (3): 430-31. doi:10.1063/1.1744152.
- Meiboom, S., et D. Gill. 1958. « Modified Spin-Echo Method for Measuring Nuclear Relaxation Times ». *Review of Scientific Instruments* 29 (8): 688-91. doi:10.1063/1.1716296.
- Mulder, Frans A. A., et Mikael Akke. 2003. « Carbonyl ¹³C Transverse Relaxation Measurements to Sample Protein Backbone Dynamics ». *Magnetic Resonance in Chemistry* 41 (10): 853-65. doi:10.1002/mrc.1252.
- Mulder, Frans A. A., Robin A. de Graaf, Robert Kaptein, et Rolf Boelens. 1998. « An Off-resonance Rotating Frame Relaxation Experiment for the Investigation of Macromolecular Dynamics Using Adiabatic Rotations ». *Journal of Magnetic Resonance* 131 (2): 351-57. doi:10.1006/jmre.1998.1380.
- Nerdal, Willy, Dennis R. Hare, et Brian R. Reid. 1989. « Solution structure of the EcoRI DNA sequence: refinement of NMR-derived distance geometry structures by NOESY spectrum back-calculations ». *Biochemistry* 28 (26): 10008-21. doi:10.1021/bi00452a020.
- Nikolova, Evgenia N., Gavin D. Bascom, Ioan Andricioaei, et Hashim M. Al-Hashimi. 2012. « Probing Sequence-specific DNA Flexibility in A-tracts and Pyrimidine-purine Steps by NMR ¹³C Relaxation and MD Simulations ». *Biochemistry* 51 (43): 8654-64. doi:10.1021/bi3009517.
- Nikolova, Evgenia N., Federico L. Gottardo, et Hashim M. Al-Hashimi. 2012. « Probing Transient Hoogsteen Hydrogen Bonds in Canonical Duplex DNA Using NMR Relaxation Dispersion and

- Single-atom Substitution ». *Journal of the American Chemical Society* 134 (8): 3667-70. doi:10.1021/ja2117816.
- Nikolova, Evgenia N., Eunae Kim, Abigail A. Wise, Patrick J. O'Brien, Ioan Andricioaei, et Hashim M. Al-Hashimi. 2011. « Transient Hoogsteen Base Pairs in Canonical Duplex DNA ». *Nature* 470 (7335): 498-502. doi:10.1038/nature09775.
- Oguey, Christophe, Nicolas Foloppe, et Brigitte Hartmann. 2010. « Understanding the Sequence-Dependence of DNA Groove Dimensions: Implications for DNA Interactions ». *PLoS ONE* 5 (12): e15931. doi:10.1371/journal.pone.0015931.
- Olson, Wilma K., Andrey A. Gorin, Xiang-Jun Lu, Lynette M. Hock, et Victor B. Zhurkin. 1998. « DNA Sequence-Dependent Deformability Deduced from protein–DNA Crystal Complexes ». *Proceedings of the National Academy of Sciences* 95 (19): 11163-68. doi:10.1073/pnas.95.19.11163.
- Palmer, Arthur G., Christopher D. Kroenke, et J. Patrick Loria. 2001. « Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules ». In *Methods in Enzymology*, 339:204-38. Elsevier. <http://linkinghub.elsevier.com/retrieve/pii/S0076687901393151>.
- Palmer, Arthur G., et Francesca Massi. 2006. « Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy ». *Chemical Reviews* 106 (5): 1700-1719. doi:10.1021/cr0404287.
- Paquet, Françoise, Florence Gaudin, et Gérard Lancelot. 1996. « Selectively ¹³C-Enriched DNA: Evidence from ¹³C1' Relaxation Rate Measurements of an Internal Dynamics Sequence Effect in the Lac Operator ». *Journal of Biomolecular NMR* 8 (3): 252-60. doi:10.1007/BF00410324.
- Pasi, Marco, John H. Maddocks, David Beveridge, Thomas C. Bishop, David A. Case, Thomas Cheatham, Pablo D. Dans, et al. 2014. « μ ABC: A Systematic Microsecond Molecular Dynamics Study of Tetranucleotide Sequence Effects in B-DNA ». *Nucleic Acids Research*, septembre, gku855. doi:10.1093/nar/gku855.
- Pelupessy, Philippe, et Elisabetta Chiarparin. 2000. « Hartmann–Hahn Polarization Transfer in Liquids: An Ideal Tool for Selective Experiments ». *Concepts in Magnetic Resonance* 12 (3): 103-24. doi:10.1002/(SICI)1099-0534(2000)12:3<103::AID-CMR1>3.0.CO;2-#.
- Pérez, Alberto, Iván Marchán, Daniel Svozil, Jiri Sponer, Thomas E. Cheatham III, Charles A. Laughton, et Modesto Orozco. 2007. « Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers ». *Biophysical Journal* 92 (11): 3817-29. doi:10.1529/biophysj.106.097782.
- Pervushin, Konstantin V., Gerhard Wider, et Kurt Wüthrich. 1998. « Single Transition-to-Single Transition Polarization Transfer (ST2-PT) in [¹⁵N,¹H]-TROSY ». *Journal of Biomolecular NMR* 12 (2): 345-48. doi:10.1023/A:1008268930690.
- Přecechtělová, Jana, Markéta L. Munzarová, Juha Vaara, Jan Novotný, Martin Dračinský, et Vladimír Sklenář. 2013. « Toward Reproducing Sequence Trends in Phosphorus Chemical Shifts for Nucleic Acids by MD/DFT Calculations ». *Journal of Chemical Theory and Computation* 9 (3): 1641-56. doi:10.1021/ct300488y.
- Sattler, Michael, Jürgen Schleucher, et Christian Griesinger. 1999. « Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients ». *Progress in Nuclear Magnetic Resonance Spectroscopy* 34 (2): 93-158. doi:10.1016/S0079-6565(98)00025-9.
- Schroeder, Steven A., Vikram Roongta, Josepha M. Fu, Claude R. Jones, et David G. Gorenstein. 1989. « Sequence-dependent variations in the phosphorus-31 NMR spectra and backbone torsional angles of wild-type and mutant lac operator fragments ». *Biochemistry* 28 (21): 8292-8303. doi:10.1021/bi00447a006.
- Schwieters, Charles D., et G. Marius Clore. 2007. « A Physical Picture of Atomic Motions within the Dickerson DNA Dodecamer in Solution Derived from Joint Ensemble Refinement against NMR and Large-Angle X-ray Scattering Data† ». *Biochemistry* 46 (5): 1152-66. doi:10.1021/bi061943x.
- Shajani, Zahra, et Gabriele Varani. 2008. « ¹³C Relaxation Studies of the DNA Target Sequence for HhaI Methyltransferase Reveal Unique Motional Properties† ». *Biochemistry* 47 (29): 7617-25. doi:10.1021/bi7020469.

- Shakked, Zippora, Gali Guerstein-Guzikevich, Miriam Eisenstein, Felix Frolow, et Dov Rabinovich. 1989. « The Conformation of the DNA Double Helix in the Crystal Is Dependent on Its Environment ». *Nature* 342 (6248): 456-60. doi:10.1038/342456a0.
- Simon, Bernd, Katia Zanier, et Michael Sattler. 2001. « A TROSY Relayed HCCH-COSY Experiment for Correlating Adenine H2/H8 Resonances in Uniformly ¹³C-Labeled RNA Molecules ». *Journal of Biomolecular NMR* 20 (2): 173-76. doi:10.1023/A:1011214914452.
- Solomon, I. 1955. « Relaxation Processes in a System of Two Spins ». *Physical Review* 99 (2): 559-65. doi:10.1103/PhysRev.99.559.
- Thåström, A., L.M. Bingham, et J. Widom. 2004. « Nucleosomal Locations of Dominant DNA Sequence Motifs for Histone–DNA Interactions and Nucleosome Positioning ». *Journal of Molecular Biology* 338 (4): 695-709. doi:10.1016/j.jmb.2004.03.032.
- Tian, Ye, Michael Kayatta, Katharine Shultis, Alejandro Gonzalez, Leonard J. Mueller, et Mary E. Hatcher. 2009. « ³¹P NMR Investigation of Backbone Dynamics in DNA Binding Sites† ». *The Journal of Physical Chemistry B* 113 (9): 2596-2603. doi:10.1021/jp711203m.
- Tidor, Bruce, Karl K. Irikura, Bernard R. Brooks, et Martin Karplus. 1983. « Dynamics of DNA Oligomers ». *Journal of Biomolecular Structure and Dynamics* 1 (1): 231-52. doi:10.1080/07391102.1983.10507437.
- Tisné, Carine, Edith Hantz, Brigitte Hartmann, et Muriel Delepierre. 1998. « Solution structure of a non-palindromic 16 base-pair DNA related to the HIV-1 κ b site: evidence for BI-BII equilibrium inducing a global dynamic curvature of the duplex1 ». *Journal of Molecular Biology* 279 (1): 127-42. doi:10.1006/jmbi.1998.1757.
- Tisné, Carine, Brigitte Hartmann, et Muriel Delepierre. 1999. « NF- κ B Binding Mechanism: A Nuclear Magnetic Resonance and Modeling Study of a GGG \rightarrow CTC Mutation ». *Biochemistry* 38 (13): 3883-94. doi:10.1021/bi982402d.
- Tjandra, Nico, Shin-ichi Tate, Akira Ono, Masatsune Kainosho, et Ad Bax. 2000. « The NMR Structure of a DNA Dodecamer in an Aqueous Dilute Liquid Crystalline Phase ». *Journal of the American Chemical Society* 122 (26): 6190-6200. doi:10.1021/ja000324n.
- Trott, Oleg, et Arthur G. Palmer III. 2002. « R1 ρ Relaxation outside of the Fast-Exchange Limit ». *Journal of Magnetic Resonance* 154 (1): 157-60. doi:10.1006/jmre.2001.2466.
- Vallurupalli, Pramodh, Guillaume Bouvignies, et Lewis E. Kay. 2012. « Studying “Invisible” Excited Protein States in Slow Exchange with a Major State Conformation ». *Journal of the American Chemical Society* 134 (19): 8148-61. doi:10.1021/ja3001419.
- Varnai, P. 2002. « α/γ Transitions in the B-DNA backbone ». *Nucleic Acids Research* 30 (24): 5398-5406. doi:10.1093/nar/gkf680.
- Vasudevan, Dileep, Eugene Y. D. Chua, et Curt A. Davey. 2010. « Crystal Structures of Nucleosome Core Particles Containing the ‘601’ Strong Positioning Sequence ». *Journal of Molecular Biology* 403 (1): 1-10. doi:10.1016/j.jmb.2010.08.039.
- Vuister, Geerten W, et Ad Bax. 1992. « Resolution enhancement and spectral editing of uniformly ¹³C-enriched proteins by homonuclear broadband ¹³C decoupling ». *Journal of Magnetic Resonance* (1969) 98 (2): 428-35. doi:10.1016/0022-2364(92)90144-V.
- Wing, Richard, Horace Drew, Tsunehiro Takano, Chris Broka, Shoji Tanaka, Keiichi Itakura, et Richard E. Dickerson. 1980. « Crystal Structure Analysis of a Complete Turn of B-DNA ». *Nature* 287 (5784): 755-58. doi:10.1038/287755a0.
- Wu, Bin, Michael Petersen, Frederic Girard, Marco Tessari, et Sybren S. Wijmenga. 2006. « Prediction of Molecular Alignment of Nucleic Acids in Aligned Media ». *Journal of Biomolecular NMR* 35 (2): 103-15. doi:10.1007/s10858-006-9004-2.
- Xue, Yi, Dawn Kellogg, Isaac J. Kimsey, Bharathwaj Sathyamoorthy, Zachary W. Stein, Mitchell McBairty, et Hashim M. Al-Hashimi. 2015. « Chapter Two - Characterizing RNA Excited States Using NMR Relaxation Dispersion ». In *Methods in Enzymology*, édité par Sarah A. Woodson and Frédéric H. T. Allain, 558:39-73. Structures of Large RNA Molecules and Their Complexes. Academic Press. <http://www.sciencedirect.com/science/article/pii/S0076687915000774>.
- Xu, Xiaoqian, Akli Ben Imeddourene, Loussiné Zargarian, Nicolas Foloppe, Olivier Mauffret, et Brigitte Hartmann. 2014. « NMR studies of DNA support the role of pre-existing minor groove variations in

- nucleosome indirect readout ». *Biochemistry*, août. doi:10.1021/bi500504y.
- Yamazaki, Toshio, Weontae Lee, Cheryl H. Arrowsmith, D. R. Muhandiram, et Lewis E. Kay. 1994. « A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of ^{15}N , ^{13}C , ^2H Labeled Proteins with High Sensitivity ». *Journal of the American Chemical Society* 116 (26): 11655-66. doi:10.1021/ja00105a005.
- Yanagi, Kazunori, Gilbert G. Privé, et Richard E. Dickerson. 1991. « Analysis of local helix geometry in three B-DNA decamers and eight dodecamers ». *Journal of Molecular Biology* 217 (1): 201-14. doi:10.1016/0022-2836(91)90620-L.
- Zgarbová, Marie, F. Javier Luque, Jiří Šponer, Thomas E. Cheatham, Michal Otyepka, et Petr Jurečka. 2013. « Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters ». *Journal of Chemical Theory and Computation* 9 (5): 2339-54. doi:10.1021/ct400154j.
- Zgarbová, Marie, Michal Otyepka, Jiří Šponer, Filip Lankaš, et Petr Jurečka. 2014. « Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA ». *Journal of Chemical Theory and Computation* 10 (8): 3177-89. doi:10.1021/ct500120v.
- Zhao, Bo, Alexandar L. Hansen, et Qi Zhang. 2014. « Characterizing Slow Chemical Exchange in Nucleic Acids by Carbon CEST and Low Spin-Lock Field R1 ρ NMR Spectroscopy ». *Journal of the American Chemical Society* 136 (1): 20-23. doi:10.1021/ja409835y.
- Zuo, Xiaobing, et David M. Tiede. 2005. « Resolving Conflicting Crystallographic and NMR Models for Solution-State DNA with Solution X-ray Diffraction ». *Journal of the American Chemical Society* 127 (1): 16-17. doi:10.1021/ja044533+.

Résumé

L'étude de la dynamique intrinsèque de l'ADN-B permet de caractériser l'espace conformationnel exploré par cette macromolécule. Cette dynamique, qui dépend de la séquence, est un facteur clé dans les mécanismes d'interaction avec les protéines, l'ADN étant plus ou moins prédisposé à s'adapter à son partenaire.

Lors de cette thèse, nous avons sondé la dynamique de l'ADN à deux échelles de temps, en nous basant sur l'étude de quatre dodécamères par Résonance Magnétique Nucléaire (RMN). Nous nous sommes d'abord intéressés à la dynamique des groupements phosphates, qui correspondent à des mouvements rapides (picoseconde-nanoseconde). Nous avons ainsi confirmé l'effet de la séquence dinucléotidique sur cette dynamique, qui peut être prédit et quantifié. Nous avons également mis en lumière pour la première fois l'étroite inter-dépendance qui existe entre déplacements chimiques du phosphore, distances internucléotides et constantes de couplage dipolaire résiduel. L'interprétation de ces observables RMN en termes de conformations des phosphates, de paramètres hélicoïdaux et de taille des sillons, montre qu'en fait ces couplages reflètent la mécanique intrinsèque de l'ADN en solution. En interfaçant ces résultats avec l'effet de séquence observé sur la dynamique des phosphates, il est aussi possible de saisir de quelle façon la conformation moyenne de la double hélice et l'espace conformationnel associé sont modulés par la séquence au niveau dinucléotidique. Enfin, des dynamiques moléculaires réalisées avec les très récents champs de force CHARMM36 et Parmbsc0_{ezOLI}, confrontées aux données expérimentales, ont permis d'apprécier le réalisme croissant des ADN simulés et ont aidé à préciser des éléments de la dynamique qui échappent à l'expérience.

Le deuxième volet de cette thèse a porté sur les mouvements de l'ADN se produisant à l'échelle de la milliseconde, encore très peu étudiés. Nous avons mis au point des expériences de dispersion-relaxation qui ont apporté la preuve de l'existence d'un échange conformationnel d'un type totalement nouveau. Cet échange ne semble apparaître que sur un type particulier de séquence, riche en A:T. Certaines régions de l'ADN, probablement spécifiques, peuvent ainsi localement évoluer vers une forme très faiblement peuplée, dont la structure détaillée reste à caractériser.

L'ensemble de ces résultats offre un panorama des capacités dynamiques de l'ADN, dépendantes de la séquence, et ouvre ainsi de nombreuses perspectives vers une meilleure compréhension des mécanismes qui guident la formation des complexes ADN-protéines.

Abstract

The study of B-DNA intrinsic dynamics enables to characterize the conformational landscape explored by this macromolecule. Indeed, binding of DNA to proteins is modulated by subtle sequence-dependent variations inherent to the dynamics of free DNA, which facilitate or disfavor the structural fit with cognate partners.

In this thesis, the DNA dynamics was investigated at two time-scales, on the basis of the Nuclear Magnetic Resonance (NMR) study of four dodecamers. First, we examined the fast dynamics (pico-nanosecond) of phosphate linkages. We confirmed that the dinucleotide sequence modulates the backbone dynamics, an effect that can be quantified and predicted. Then, our experimental data enabled to establish that phosphorus chemical shifts, internucleotide distances and residual dipolar couplings constants are closely correlated. The translation of the NMR observables in terms of phosphate conformations, helicoidal parameters and minor groove dimension, allowed the structural interpretation of the couplings and led to the first coherent description of the intrinsic DNA mechanics in solution. Owing our knowledge of the effect of the sequence on the backbone behavior, it is now possible to understand how the DNA shape and the associated conformational landscape are modulated at the dinucleotide level. Finally, the performance of molecular dynamics (MD) simulations with the recent force-fields Parmbsc0_{ezOLI} and CHARMM36 was tested extensively against our NMR data. We found impressive progress towards a realistic representation of DNA, despite residual shortcomings. This advance allowed to reveal new aspects of the DNA dynamics, which cannot be assessed from experiments.

The second part of this thesis focused on slow motions in B-DNA, which are still largely under-investigated. Using and developing sophisticated relaxation-dispersion NMR experiments, we demonstrated the existence of a new conformational exchange at the millisecond time-scale, which seems to only occur in a particular type of sequence, A:T rich. Thus, in addition to the familiar structural patterns that are the signature of the B double helix, some short DNA regions, likely specific, are able to explore another conformational state, weakly populated, whose detailed structure still needs to be characterized.

Overall, these results provide original insights on the DNA dynamic repertoire, sequence-dependent, and open the way towards a better understanding of the mechanisms underlying the formation of DNA-protein complexes.